Evaluating Fault Management Operations Concepts for Next-Generation Spacecraft: What Eye Movements Tell Us

Intelligent Spacecraft Interface Systems (ISIS) Lab
Human Systems Integration Division &
Intelligent Systems Division
NASA Ames Research Center
Moffett Field, CA 94035

November 6, 2009

Contributors

Miwa Hayashi, Ph.D.
Ujwala Ravinder, M.S.
San Jose State University
Moffett Field, CA

Robert S. McCann, Ph.D.
Brent Beutter, Ph.D.
Lily Spirkovska, Ph.D.
NASA Ames Research Center
Moffett Field, CA

Acknowledgments

This work was supported by the Space Human Factors and Habitability Element (WBS 466199) of NASA's Human Research Program, the ARC Crew Cabin 1% ESMD AUG (WBS 64423.02.35.02.99), the ARMD Aviation Safety Program (IVHM Project), and the ARMD NRA -- Aviation Safety: Integrated Vehicle Health Management Project, Solicitation #NNH06ZEA001N-IVHM.

We would like to express our appreciation to Lee Stone for his helpful comments on the eye-movement analyses. We also thank Fritz Renema, Steven Elkins, David Garcia, Scott Poll, Adam Sweet, William Taylor, John Ossenfort, Matthew Daigle, Indranil Roychoudhury, Ann Patterson-Hine, Serdar Uckun, David Nishikawa, David Hall, Charles Lee, Christian Neukom, and Serge Yentus for their hardware and software support. Last but not least, we are grateful to our study participants for being patient with us throughout the training and data collection sessions.

Executive Summary

Real-time fault management during ascent and entry ranks among the most safety-critical of spacecraft operations. On the shuttles, confusing Caution and Warning system indications, outdated cockpit display formats, and "hard" (manual) fault isolation and recovery interfaces exacerbate fault management time and difficulty. On next-generation spacecraft, these "legacy" display formats will be replaced by state-of-the-art display formats and "soft" (electronic) crew-vehicle interfaces. Previous studies indicate that, relative to shuttle, these interfaces will enhance crewmembers' fault management performance considerably. Opportunities exist to enhance performance even further by incorporating additional forms of fault-management automation, particularly in the areas of fault diagnosis and information retrieval. However, next-generation vehicles face very stringent weight, schedule, and cost constraints. Ideally, vehicle designers should have access to quantifiable metrics on the performance benefits provided by candidate forms of automation to make informed cost/benefit analyses.

Performance enhancements associated with selected forms of automation were quantified in a recent human-in-the-loop evaluation of two candidate operational concepts for fault management on next-generation vehicles. The baseline concept, called Elsie, featured a full-suite of "soft" fault management interfaces. In many ways, fault management with Elsie was more advanced than on the shuttles (for example, off-nominal procedures were worked through an electronic procedure viewer [EPV] linked to virtual switch panels). However, operators were forced to diagnose malfunctions with minimal assistance from the standalone caution and warning system, which deliberately emulated the workings (and limitations) of the caution and warning system on today's shuttle.

The other concept, called Besi, incorporated a more capable C&W system with an automated fault diagnosis capability. In addition, Besi included functional links between the C&W database of fault messages and the EPV database of off-nominal checklist titles, which automated the process of bringing up the correct procedure checklist on the EPV. In exchange for these and other targeted investments in fault management software, Besi provided a more streamlined, user-centered fault management concept than Elsie.

Eight trained participants worked one or more systems malfunctions during simulated Orion ascents, some with Elsie and some with Besi. In parallel with their fault management duties, operators were tasked with noticing and responding to occasional color changes on their primary flight display (PFD) symbology. Operators worked systems malfunctions more quickly, accurately, and with less reported workload with Besi than with Elsie, particularly when they had to manage more than one malfunction. In addition, they missed fewer color changes on the PFD.

These results were summarized in an earlier report (Hayashi, McCann, Beutter, Spirkovska, Poll, & Sweet, 2007). Since then, additional analyses were completed of operators' manual inputs and eye movement behavior. The results more precisely resolve the source of Elsie's advantages, and provide better understanding of operators' display usage, information acquisition, and multi-tasking strategies. The most important findings, together with their design implications, are as follows:

1. On average, Besi's automated fault diagnosis capability saved approximately 19 sec of fault management time. Automating the process of bringing up the target checklist saved an additional 20 sec. The software required to functionally link the advanced C&W system with the EPV database was minimal compared to the software required for the automated diagnosis capability; thus, straightforward information integration provided the biggest operational "bang" for the software investment "buck".

2. For both Elsie and Besi, most of the time required to complete proximal ("root")-cause fault diagnoses and bring up the designed off-nominal checklist on the EPV was consumed by processing text on cluttered (text-rich) display formats. With Elsie, however, operators incorporated extensive processing of fault messages on a C&W fault log display into their fault diagnosis activities, and had to navigate through one or more text-based menus to bring up the correct checklist of fault management procedures. Processing text from these sources consumed an average of 31 seconds for Elsie versus just 11 seconds for Besi. These findings suggest that fault management system designers should seek to minimize operators' reliance on text processing wherever possible, particularly on high-density text-rich displays. Task elements that require text processing should be high-priority targets for automation. If automation is not possible, high-density displays should be structured to reduce search-based behavior.

3. Analyses of fixation patterns and transition probabilities between system summary displays, C&W fault messages, and the EPV revealed that operators processed individual sources of fault management information repeatedly, often after consulting other concurrent sources of fault-management information (cross-checking). The pattern strongly suggests that if vehicle designers opt for a more Elsie-like (less computationally demanding) concept of fault management operations, C&W system interfaces, system summary displays, and the EPV should not be forced to time-share design real estate; instead, they should be viewable simultaneously, possibly on a single consolidated fault-management display format. Time-sharing these information sources would be much less detrimental with a more advanced (Besi-like) concept.

4. During the time that operators were actively working a malfunction, they regularly interrupted their fault management activities to check the Primary Flight Display. If the malfunction was being worked with more machine assistance (Besi), these checks were performed more frequently than when more of the fault management burden was on the operator (Elsie). The increased willingness to interrupt fault management and redirect resources to the concurrent task is evidence that in a demanding multi-tasking environment, like ascent/entry, high-level strategies for time-sharing information acquisition activities across concurrent tasks are sensitive to the level of automated support. Given that these high-level (strategic) adjustments to information sampling behavior were accompanied by better performance on the PFD-based color-noticing task, automating high-workload components of fault management benefited all tasks in the environment. This result should factor into cost/benefit decision-making concerning the efficacy of automated support tools in high-workload multi-tasking environments like next-generation spacecraft ascents and entries.

Table of Contents

1. Background

1.1. Fault Management Operations on Shuttle

Suppose an onboard system experiences a malfunction during the most dynamic phases of a spacecraft mission, ascent or entry. Crewmembers must be alerted immediately so they can diagnose and understand the source of the problem, and then complete a checklist of procedures to minimize operational impacts and (if possible) maintain (or restore) mission- and safety-critical aspects of system operation. These activities often have to be completed quickly to prevent the problem from escalating into a life- or mission-threatening situation. Along with the extreme time pressure, there is little safety margin for a crew error, such as misdiagnosing the source of the problem or executing an incorrect isolation and recovery procedure.

To meet the need for fast and accurate fault management, astronauts spend hours practicing fault management-related operations in ground-based simulators. In the case of the shuttles, these training requirements are greatly exacerbated by the fault management interfaces. Apart from sounding cockpit alarms, the caution & warning (C&W) system, designed in the 1970's, does little more than flag off-nominal sensor values, generate a visual indicator (typically, an up or down arrow) beside the out-of-limits value on the appropriate system summary display, and generate fault messages. Because the shuttle systems are so complex and highly interconnected, a failure of one component often generates off-nominal operational modes, and out-of-limits sensor readings, in subsystems and equipment located "downstream" of the initiating malfunction. All too frequently, the result is a cascade of C&W alarms, multiple fault indications across several system summary displays, and a lengthy list of fault messages on the "fault log" display.

Collectively, these cockpit indications constitute a C&W system "event". For the crew, a C&W event forms a collection of often confusing symptoms that must be carefully evaluated in order to determine the source of the problem - what is sometimes labeled the "parent" malfunction. Functionally, this diagnostic process usually culminates in choosing the fault message judged to be most closely associated with (or most proximal to) the parent. This is because fault messages are isomorphic with the titles of the off-nominal checklists in the onboard flight data files (or cue cards), and so the checklist is selected by matching the parent fault message to a checklist title.

Once the crew has selected the appropriate checklist, operational challenges continue through the fault isolation and recovery procedures. Navigating through paper checklists is a complicated, multi-tasking activity that requires frequent switching of attention between distinct information sources in support of multiple sub goals and activities. For example, a common procedure calls for the operator to reconfigure the operational mode of the system experiencing the malfunction (e.g., by opening a flow control valve that is normally closed). To make it so, the crewmember must locate and physically toggle a switch on one of the many switch panels that cover the interior of the shuttle cockpit. Once the switch is thrown, the crewmember must verify that the new operational mode has been achieved, either by examining "talkback" indicators on the switch panel itself, or by checking sensor readings on a systems summary display. He (or she) must then shift attention back to the checklist and re-establish (from memory) what steps have been completed, and what the next to-be-completed step is.

Adding to the difficulty of checklist navigation, checklists typically contain logical conditionals (e.g., “If nonisolatable” or “If isolated”) that, depending on their resolution, establish different pathways through the remainder of the list, with some (to-be-performed) procedures staying on the pathway, and others (to-be-ignored) falling off. With few or no visual cues (e.g., arrows, lines, common indentation levels, etc.) to link the individual procedures of a pathway together, the operator must construct and maintain a mental representation of the pathway – and keep track of his or her place within it – mostly from memory. In a recent study of fault management performance in a part-task shuttle simulator at NASA Ames Research Center (McCann, Beutter, Matessa, McCandless, Spirkovska, Liston, Hayashi, Ravinder, Elkins, Renema, Lawrence, & Hamilton, 2006), researchers found that checklist completion times lengthened dramatically when the pathway encompassed widely separated procedures. In one case, involving a safety-critical leak in the helium supply system to the shuttle’s main engines, several participants failed to navigate to the final “orphan” procedure at all, despite extensive training that emphasized it.

As if all these difficulties were not enough, during the dynamic flight phases of ascent and entry, the crew must time-share any fault management activities that may arise with other critical tasks, such as monitoring the vehicle’s attitude, velocity, and flight path on their primary flight display. Checklist developers endeavor to relieve the burden during these periods with "short-list" procedures that postpone as many activities as possible to a less dynamic period, such as when the shuttle reaches orbit. Unfortunately, limitations in sensor coverage and other factors often render the exact nature or location of a malfunction ambiguous. In these cases, even the short lists must include preliminary troubleshooting procedures to determine the nature of the malfunction more precisely, adding considerably to malfunction resolution time and crew workload.

In summary, fault management on the shuttles – with their confusing C&W interfaces, legacy system summary display designs, and requirements for paper checklist navigation – can easily overwhelm the crews’ attentional and cognitive processing resources, particularly during the dynamic phases of flight when these resources are already stretched thin. Fortunately, the shuttles operate close enough to the Earth that communications systems are able to provide a near-real-time stream of vehicle telemetry (including sensor readings) to Mission Control Center (MCC). Along with the crew, MCC flight controllers and systems subject-matter experts continuously monitor these data for indications of anomalous behavior. Much like the onboard C&W system, ground software flags out-of-limit parameters, sometimes using tighter limits than those used onboard in order to detect potential faults more quickly. Additionally, as ground software can be upgraded more easily than Shuttle software, it is closer to the state-of-the-art in sensor fusion and data processing capabilities. For example, ground software provides trending information on some main engine parameters in graphical form. Armed with these tools, ground personnel typically assist the crew with disambiguating the parent of a C&W event and determining the appropriate response. A crewmember must still locate and flip the proper switches and verify that isolation and recovery procedures are proceeding as expected.

1.2. Fault Management on Next-Generation Vehicles: Challenges and Opportunities

Decades ago, an ongoing revolution in aircraft operations was initiated by the arrival of the first generation of glass-cockpit aircraft, featuring integrated avionics architectures and soft (electronic) crew-vehicle operational interfaces. Although the shuttle cockpits were upgraded from the original electromechanical instruments and CRT displays to glass starting in the late 1990s, the display formats (and associated operations concepts) were largely ported from the original versions. Fully recognizing the opportunities afforded by glass, shuttle operations experts completed a comprehensive cockpit avionics upgrade (CAU) project, including a complete redesign of the shuttle cockpit displays (McCandless, Hilty, and McCann, 2005). Due to time and budget constraints, the shuttle cockpits were never upgraded to support the redesigned displays. However, a high fidelity shuttle mission simulator at NASA Johnson Space Center was upgraded to support the CAU display suite, and a thorough human-in-the-loop evaluation of the operational impact of the redesigns was completed in that facility. "Crews" assembled from astronaut office personnel worked a wide variety of shuttle systems malfunctions during short periods of simulated ascent and/or entry, once with the existing display formats and again several months later with the upgraded formats. With the existing display suite, crewmembers failed to recognize (diagnose) fully 30% of the problems in the highest workload scenarios, where multiple independent malfunctions occurred in close temporal proximity. Only 10% of these malfunctions were unrecognized with the upgraded suite.

By observing the crews in real time, subject matter experts were also able to assess the impact of the redesigned display suite on how *long* it took crewmembers to diagnose malfunctions. The subset of malfunctions whose diagnosis times placed them in the slowest quartile, relative to the entire set of malfunctions included in the study, took an average of almost two minutes to diagnose with the current display suite (presumably because these malfunctions were associated with the most confusing C&W events). The average diagnosis time was reduced to 76 seconds with the upgraded display suite, for a full 44-second reduction.

The fact that the redesigned displays yielded sizable reductions in the time to understand the malfunctions is noteworthy. The longer a system remains in a faulty state, the more likely it is to degrade to the point where functionality cannot be restored. For instance, the longer an auxiliary power unit runs with an oil leak, the more likely it is that the reservoir will empty and the unit will seize (or worse). Less obviously, the faster a crewmember can work a malfunction, the lower the chances that a second, unrelated malfunction will occur before he or she has finished with the earlier problem. Human operators have very limited capacity to handle even one malfunction in conjunction with the other operational demands they face during dynamic flight; having to handle more than one degrades their performance considerably (McCann, et al., 2006).

The results of the CAU evaluation were clear: modern, task-oriented system summary displays produce dramatic improvements in the crew' ability to make sense of their cockpit indications and maintain better situation awareness of system state and status. However, the CAU redesigns were limited to existing electronic displays and interfaces; the CAU project did not attempt to replace the hard crew-vehicle interfaces in the shuttle cockpit, such as the paper-based procedure booklets and hard switches, with electronic versions. The new generation of Project Constellation vehicles will have much less interior room than the shuttles, and are being designed amid stronger pressures to minimize vehicle weight and development costs. Consequently, virtually all vehicle operations will take place through a small number of electronic (glass) panels. As far as fault management is concerned, an electronic procedure viewer (EPV) will replace the paper checklists found on the shuttles, and operators will perform mode reconfigurations largely via "soft" (electronic) representations of switches (or switch functions), rather than hard (physical) switches.

The transition to a full suite of "soft" crew-vehicle interfaces presents operational concept designers with abundant opportunities to improve fault management operations. For example, following the lead of EPV designs on modern glass cockpit aircraft, the preliminary design for the Orion Crew Exploration Vehicle EPV includes a colored focus bar that highlights the next to-be-completed line in the fault management checklist. The bar moves automatically to the next line in the procedure when the current procedure is complete, automatically skipping lines that fall out of the pathway. Through features such as these, the EPV will simplifies the task of navigating through a checklist and lessen the navigational workload on the operator (McCann et al., 2006).

Of course, there is no such thing as a free lunch. All information displayed on soft interfaces has to be represented electronically. Useful as they are to the operator, advanced display features, such as the EPV focus bar, obviously require software. Compared to the software requirements for fault management on shuttle, therefore, the wholesale conversion from hard to soft interfaces on next-generation vehicles will entail unavoidable increases in software development, testing, and verification requirements, and increase requirements for onboard computing resources and computer memory.

As we've noted, designers are facing strong pressures to keep the software and hardware requirements associated with next-generation vehicles to a minimum. Since the transition to soft cockpit interfaces is virtually mandated, so are the additional software and hardware requirements that accompany them. However, any effort to improve fault management performance with additional forms of automation, that require *additional* software, will be scrutinized very carefully, with the bar set quite high for acceptance.

1.3. Additional Opportunities

In fact, operational elements that might benefit from additional automation are not hard to identify. In their assessment of cockpit-related problems with current shuttle operations, CAU team members called out the confusing and overwhelming nature of the C&W events as one of the biggest human factors problems with the cockpit. Accordingly, a team of C&W system experts developed an enhanced caution and warning (ECW) system that applied straightforward "rules of thumb" to the fault messages generated by the C&W system in order to identify the message most closely associated with the parent malfunction. Combined with a well thought-out concept for crew-ECW interfaces (such as a scheme to suppress "daughter" fault messages and other alerts), the automated fault diagnosis capability of ECW held great promise for improving both the efficiency and the accuracy of fault diagnoses on shuttle, over and above the improvements associated with the redesigned CAU display formats alone. However, the ECW system was never implemented, even in ground-based shuttle simulators, and was therefore not included in the ground-based CAU evaluation. Although hard numbers on the operational benefits associated with automating the fault diagnosis process were never obtained, an advanced C&W system with automated diagnostic capability is clearly one of the top candidates for Orion.

The other obvious candidates for targeted automation are operations that can benefit from the fact that, for the first time, all forms of fault management information will be represented electronically. Electronic representation provides designers with an opportunity to engineer a functionally integrated software architecture that automates the process of accessing and bringing up needed fault management-related forms of information. A specific example from modern glass-cockpit aircraft is as follows. In most cases, the fault messages in the C&W system database are isomorphic with the off-nominal checklist titles in the paper ascent/entry systems procedures flight data file. To select the correct checklist of off-nominal procedures, operators must match their choice of the proximal-cause fault message to the corresponding title in the appropriate cue card or flight data file. Just as it still is on shuttle, this activity used to be quite manually intensive, often involving flipping through several pages of checklists before locating the matching title. By functionally linking the electronic database of fault messages with the electronic database of EPV checklists, however, designers of the B777 enabled operators bring up the target off-nominal checklist by simply selecting (clicking on) their choice of proximal-cause fault message from the list generated by the C&W system.

1.4. Problem Statement

Fault management operations on Orion stand to benefit greatly from the full suit of modern electronic crew-vehicle interfaces. Additional benefits may be achieved with investments in targeted forms of fault-management automation, such as an advanced C&W system with automated fault diagnosis capabilities and a unified avionics system that functionally links historically isolated repositories of fault management information. However, these investments would impact software development, testing, and verification schedules, requirements for onboard computing hardware, and vehicle weight. The outstanding question is whether the operational enhancements that accompany these investments are sufficient to justify the cost, schedule, and weight-related impacts. In order to make that determination, vehicle designers must have quantitative measures of the operational benefits that accompany the automation.

Such metrics are not available from previous studies. The appropriate baseline against which to assess automation benefits must include a full suite of electronic interfaces, which the CAU upgrade evaluation did not include. What is needed is a direct, quantitative comparison of fault management between two operations concepts, both of which employ a full suite of soft crew-vehicle interfaces, but only one of which includes the additional sources of targeted automation.

1.5. The Present Study

In a preliminary effort to assess these impacts, we recently completed a human factors evaluation of two fault management concepts for next-generation spacecraft in the Intelligent Spacecraft Interface Systems (ISIS) Lab. The less computationally demanding concept, called Elsie, coupled a full suite of next-generation "soft" (electronic) interfaces, including an EPV, with a "bare-bones" avionics architecture featuring a stand-alone caution and warning (C&W) system patterned after the current C&W system on shuttle.

Elsie's electronic interfaces incorporated many features designed to streamline and enhance fault management activities, most of which are not available on hard paper checklists and physical switch panels. However, Elsie's shuttle-like C&W system featured very limited capabilities and had no links to other sources of fault-management information. These limitations established Elsie as representative of a next-generation fault-management concept with minimal software development, testing, integration, and certification requirements and minimal onboard hardware requirements. In return, however, Elsie retained many of the same fault management difficulties that operators encounter on shuttle. Elsie's C&W system typically generated a confusing cascade of alarms, fault messages, and off-nominal parameter indications in response to a malfunction. Operators were forced to work through these symptoms to diagnose the parent ("proximal cause") of the C&W event, a time consuming and error prone activity for even very highly trained individuals. In addition, the lack of any avionics integration forced crewmembers to access needed information, such as the checklists of fault management procedures, via look-up menus. Notably, this method of checklist retrieval may represent a hidden cost of going electronic compared to shuttle, where cue cards containing the checklists for the most critical ascent and entry malfunctions are velcroed to the cockpit, rendering them easily available to operators without having to flip through pages of flight data files. In Elsie, on the other hand, operators had to manually navigate through an electronic menu of checklist titles to locate and reach all checklists.

Elsie provided an appropriate baseline condition against which to quantify the performance benefits that would accompany our second, more advanced concept, called "Besi". Besi boasted a more capable C&W system than Elsie, complete with automated proximal (root) cause diagnoses for C&W system events. In addition, in Besi, the failure messages generated by the C&W system were dynamically linked to the EPV database of off-nominal checklist titles. This linkage allowed operators to bypass the process of navigating through EPV menus to bring up the target checklist. Instead, operators had the option of bringing up a procedures checklist by simply selecting the fault message selected by the C&W system as the proximal cause of the C&W event.

Our empirical evaluation compared operators' performance with Elsie and Besi over a series of simulated Orion ascents. Each ascent included one or two independent malfunctions in a simulated electrical power system (EPS) modeled after the Advanced Diagnostics And Prognostics Technologies (ADAPT) hardware testbed at NASA Ames Research Center (Poll, Patterson-Hine, Camisa, Garcia, Hall, Lee, Mengshoel, Neukom, Nishikawa, Ossenfort, Sweet, Yentus, Roychoudhury, Daigle, Biswas, and Koutsoukos, 2007; see Hayashi et. al., 2007, for additional details). Operators had to detect, diagnose, and respond to the malfunctions by selecting and completing the appropriate checklist of fault isolation and recovery procedures.

Along with these fault management activities, operators were responsible for detecting occasional changes in the display color of one of a key set of PFD flight parameters, such as the g-meter, thrust indicator, or vehicle velocity. Upon noticing a color change, the task was to reach out and physically touch the PFD parameter exhibiting the color change, and verbally annunciate that parameter's name. By including this continuous monitoring task, we were able to evaluate Elsi/Bessi performance differences in a multitasking environment commensurate with the environment facing vehicle operators during dynamic phases of spacecraft flight.

The previous report on this study (Hayashi et al., 2007) summarized the results of the evaluation on a standard suite of human factors performance metrics, such as the accuracy with which operators diagnosed the parent of a C&W event, the time they took to complete critical sub elements of the fault management process, and their subjective workload ratings. As expected, Besi typically supported better performance on these measures. On some malfunctions, Besi reduced diagnosis time from over 40 seconds to less than 25 seconds. Besi also supported significantly more accurate fault management performance (an amalgamation of diagnosis and procedures completion accuracy) on multi-malfunction trials, when the information processing demands on the operator were highest. Compared to Elsie, Besi yielded lower subjective workload ratings, more so on multi-malfunction trials than on single malfunction trials. That is, the higher the fault-management-related workload on the operators, the more benefit Besi provided.

Last but not least, there were preliminary indications that Besi supported a more effective division of operators' attention across the fault management task and PFD-based color detection task. During the time that operators were actively working a malfunction, they failed to respond to 30 color changes on the PFD while working malfunctions with Elsie. They missed only three color changes with Besi.

1.5.1. Limitations of Traditional Measures

Human factors evaluations of spacecraft operational concepts are typically performed after the operations concept (and associated interfaces) is relatively mature, often to determine whether the concept meets top-down guidelines for aggregate measures of performance such as workload and error rate. From the perspective of a spacecraft designer, however, a much more valuable evaluation product would be a set of specific guidelines and recommendations for the design of an operational concept and its supporting interfaces. Such guidance requires a more fine-grained analysis of operators' information acquisition strategies and display format usage than traditional measures of performance are able to provide. What displays caused the operators the most trouble, in terms of time? What forms of automation in Besi were the most germane to producing the aggregate performance benefits? What forms of automation produced the most benefit for the least software requirements delta?

This point is worth expanding. In the multi-tasking environment of a spacecraft cockpit during dynamic flight, many sources of information compete for the operator's attention. Understanding how (and for how long) the operator focuses his or her attention on a particular source can provide important guidance for designers. For example, if a source is available but ignored, either the operators aren't following the designer's model for how they would perform the task, or the information source is fully redundant with other sources, which operators prefer to use.

In the case of displays that are forced to time-share display real estate, for example, an upper limit on how long a display was available for viewing is available from time stamps of display-navigation-related button presses, which lock in when one display was swapped out in favor of another. However, such results aren't definitive, as it's not clear how much time the operator was actually viewing that display, as opposed to other regions of interest simultaneously available.

Precise measurements of the proportion of time that a particular information source was being consulted are available only through analyses of operators' oculomotor behavior (eye movements).

Accordingly, throughout each simulated ascent, operator's eye movements were recorded continuously by an ISCAN eye-tracking system. Considerable post-processing was necessary to analyze and interpret eye movement data (see Section 2 below), and most of this activity was outside the short time frame required to produce the earlier report. A cursory analysis of eye movement data did, however, suggest a preliminary connection between the number of missed color changes on the PFD, and operator's scanning behavior. The display was split into upper (PFD) and lower (malfunction handling) regions, and the percentage of time observers spent looking at the PFD versus the lower fault management region was calculated during each active malfunction-handling period. The results showed that operators looked at the PFD approximately 30% of the time when working malfunctions with Besi, compared to 24% with Elsie. This difference was statistically significant.

Once the oculomotor data were processed appropriately, the actual spatial resolution obtained supported a division of the lower (fault management) display region into several functionally distinct sub regions. Analyses of operators' information acquisition activity with respect to these sub regions form the basis of the rest of this report.

2. Methodology

2.1. Simulator Facility

The experiment was conducted in the CEV Orion simulator at the Intelligent Spacecraft Interface Systems (ISIS) laboratory at NASA Ames Research Center (Figure 2-1). The top half of the 20-inch touch-sensitive monitor on the left side displayed the Primary Flight Display (PFD), and the bottom half presented the ACAWS display. (The monitor on the right side was not used in this study.) The operators used the Nostromo hand controller (Figure 2-2) with an orange button to silence alarms, a castle switch to control and move cursor focus, and a "select button" to click on selectable display elements, such as edge key labels. A speaker system was installed in the simulator room to provide spacecraft engine noise and auditory alarms. Throughout the study, operators wore a baseball cap outfitted with an eye-tracking system (ISCAN ETL-500 eye tracker integrated with Polhemus FasTRAK head tracker), which computed the participant's gaze point on the monitor with up to 60 Hz of temporal resolution, and approximately 0.5 inch of spatial resolution.

2.1.1. Displays and Procedures

The following sections describe and explain the Elsie and Besi fault management concepts. In general, Elsie was more similar to shuttle in terms of the level of automation assistance with fault management operations, with exceptions of the electronic switches and EPV, while Besi

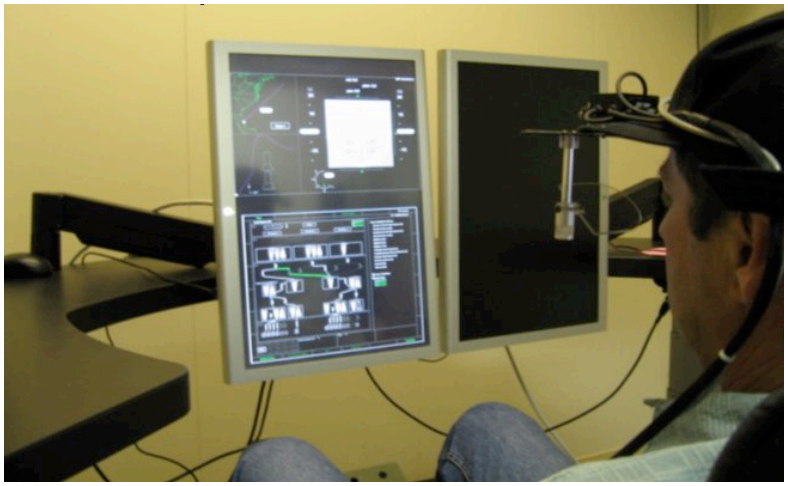


Figure 2-1. CEV Orion Simulator

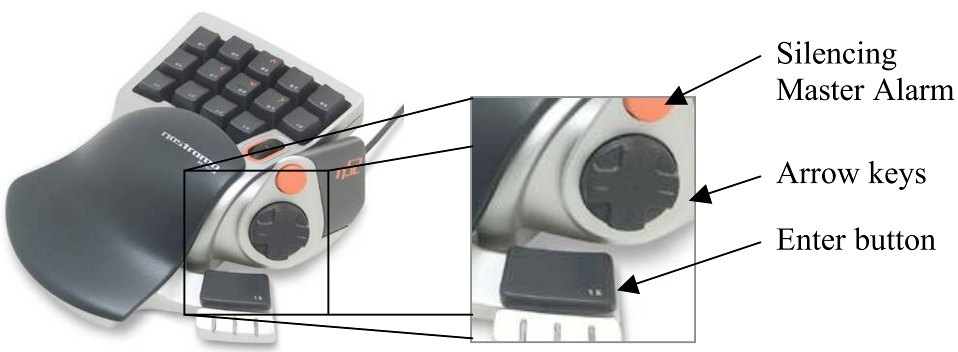


Figure 2-2. Nostromo Hand Controller

provided additional forms of automated assistance. Display formats relating to the electrical power system (EPS) and the PFD were developed specifically for this study, while the other displays were quite similar to display formats developed as part of the CAU project.

2.2. Elsie

Figure 2-3 shows the fault management interfaces for Elsie. The interface consisted of three sub regions: the Main Area, the Checklist Area, and the Fault Message Area. In Figure 2-3, the Main Area is showing the EPS Summary (EPS Sum) display, a line-diagram of the EPS. The main concept behind the Elsie design was to show the crew detailed system information so that s/he could make informed diagnosis of the proximal cause of C&W events. System information was provided on several Main Area displays: EPS Sum, which showed the EPS line-diagram (Figure 2-3); EPS Main, which provided all available EPS parameters in a table format (Figure 2-4), and EPS Loads, which indicated the status of the load switches in a table format (Figure 2-5). All ECLSS parameters were presented in the ECLSS display (Figure 2-6). These displays all appeared in the Main Area and could be called up via dedicated edge key labels along the top of the screen. The virtual switch panels (Figure 2-7 shows an example) were also presented in the

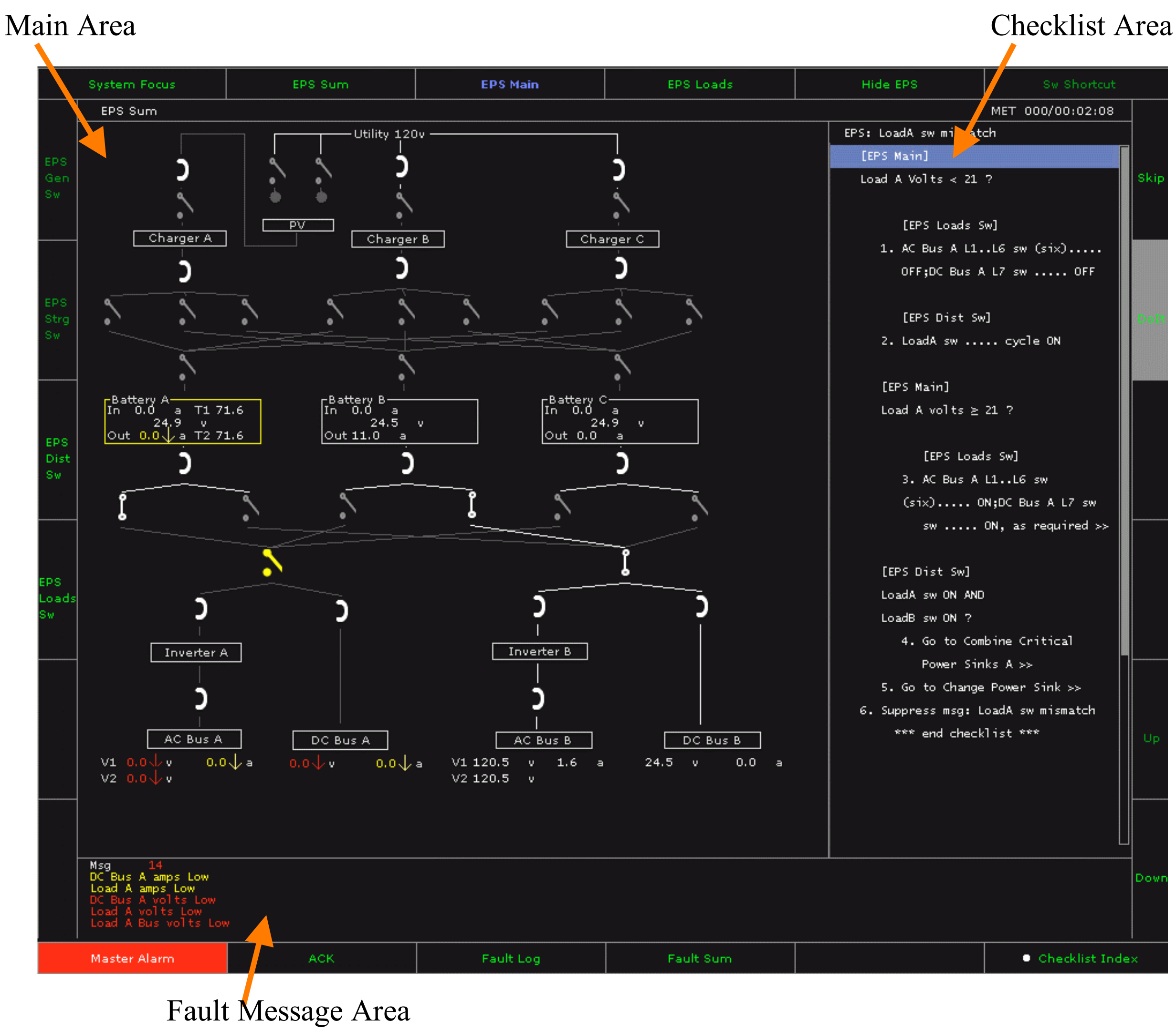


Figure 2-3. Elsie

Main Area, and the green box (cursor focus area) could be moved among the switch icons (Figure 2-7).

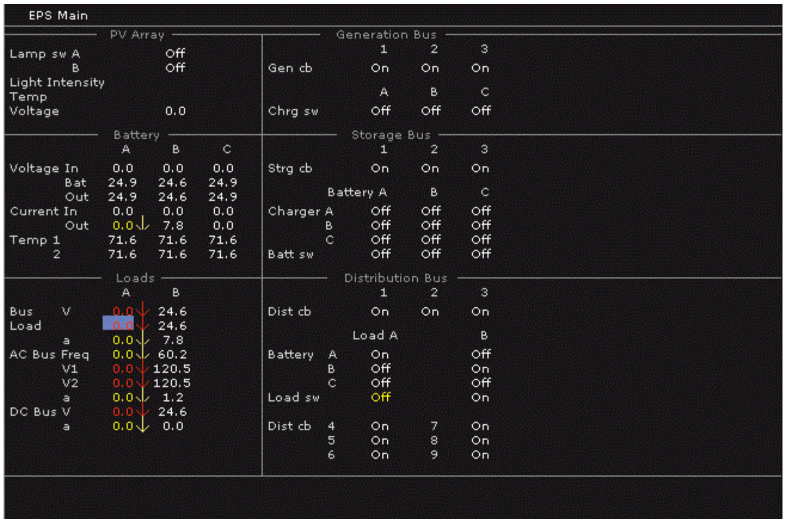


Figure 2-4. Elsie - EPS Main

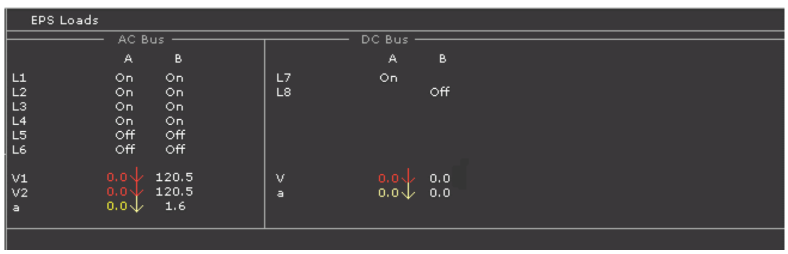


Figure 2-5. Elsie - EPS Loads

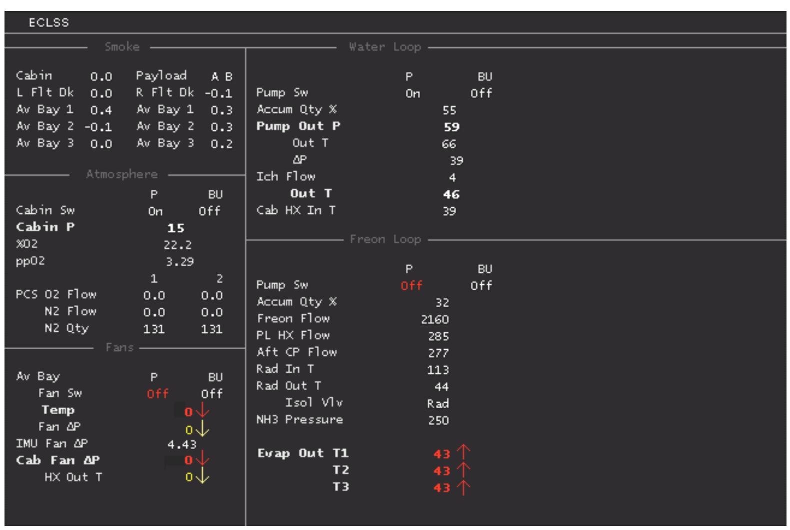


Figure 2-6. Elsie - ECLSS

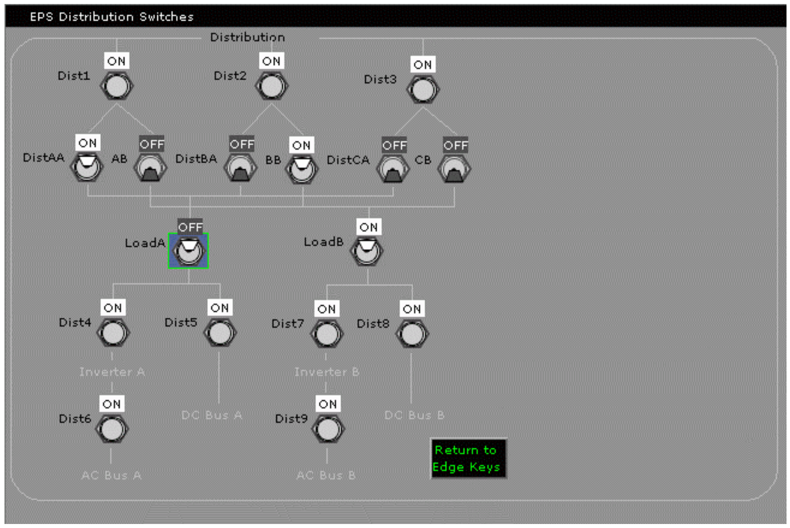


Figure 2-7. Elsie - EPS Distribution Switch Panel

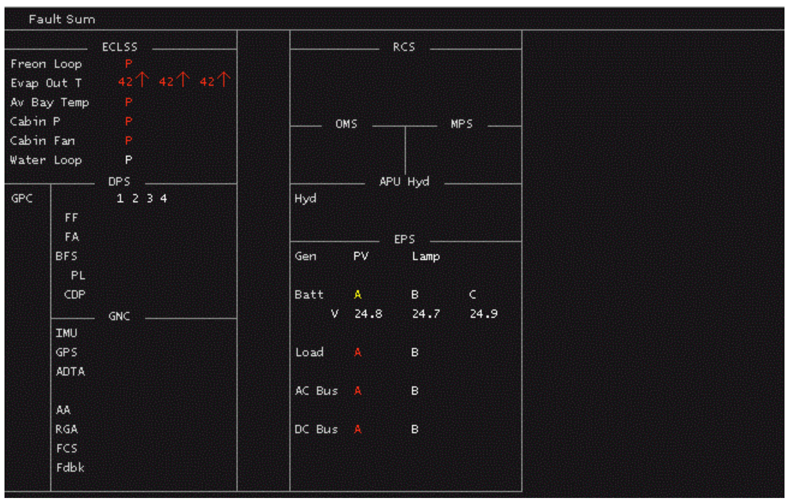


Figure 2-8. Elsie - Fault Sum

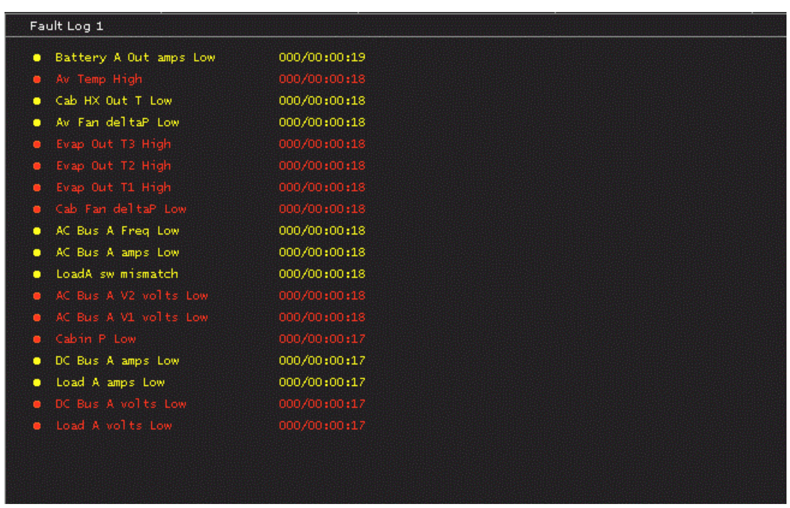


Figure 2-9. Elsie - Fault Log

Two additional displays that could be selected and viewed in the Main Area were the Fault Summary display (Fault Sum; Figure 2-8) and Fault Log (Figure 2-9), which contained the full list of fault messages generated by the C&W system. The typical fault management process unfolded as follows. Prior to the insertion of a systems malfunction, participants were instructed to divide their attention between Fault Sum, which was the default display for the Main Area, and the PFD. Fault Sum was designed to provide "at a glance" health status of the EPS and ECLSS systems. When a malfunction occurred, the C&W system responded with cockpit indicators that collectively formed the C&W event: a visual alarm (the red Master Alarm edge key label at the lower left of Figure 2-3); an auditory alarm; changes in the display color of the values and symbols associated with the affected component(s) on Fault Sum (yellow for caution [affecting non-critical loads], red for warning [affecting critical loads]); and up to five color-coded fault messages in the lower Message Area. Operators were instructed to start utilizing the color-coding pattern on Fault Sum to assess which system, EPS or ECLSS, was likely to contain the proximal cause of the C&W event. They were then encouraged (though this was optional), to check the EPS Sum display (if the suspected component was in EPS) or the ECLSS display (if it was in ECLSS) to further evaluate which specific component represented the most likely proximal-cause failure.

In addition to these graphical displays, the fault messages issued by the C&W system provided an additional and important source of diagnostic information. For its part, the Message Area had room enough to display only the last five messages generated (i.e., the fault message area was populated on a on a "last-in" basis). If (as was typically the case) the fault generated more than five messages, the operator could only view the full set by bringing up the Fault Log display (Figure 2-9) in the Main Area. Fault Log had three pages, each of which could list up to eighteen messages, or 54 in total, in chronological order (with the most recently generated [newest] message at the top). As a "rule-of-thumb" for diagnosing the parent malfunction, operators were instructed to look for any "switch mismatch" fault message (i.e., the commanded position and sensed position of the switch were in disagreement). If there was no such message, operators were instructed to then search for a volts-related message from the most upstream component of the EPS/ECLSS hierarchy.

Fault messages also played a pivotal role in the second phase of the fault management operations, labeled C1, which was to navigate to and bring up the appropriate checklist in the EPV. Critically, the labels populating the menus of EPS checklists were identical to a C&W fault messages. In the case of Elsie, therefore, operators had to first select the one message judged to correspond to the proximal cause of the C&W event, and then match that message to an entry in the EPV menu.

Once the diagnosis was completed, the next step was to locate (navigate to) and select the corresponding label from the menu of EPS malfunction checklists in the EPV (see below). Functionally, the operator accomplished these activities by moving the cursor focus to the *Checklist Index* edge key label, in the lower right corner, and selecting (clicking on) it. Clicking *Checklist Index* brought up the menu of EPS failure checklists in the EPV area (in practice, it was not uncommon for a participant to bring up the checklist menu before making a final selection from the fault log messages, suggesting operators sometimes engaged in a pattern-matching activity between fault messages and checklist labels which may have helped them with the diagnosis itself). Then, the operator manually navigated to the target checklist label and selected it, thereby bringing up the target checklist on the EPV. Once the target checklist was accessed, operators used edge key labels on the right side of the fault management display to navigate through (and complete) the individual procedures (lines).

In summary, these message-related activities in Elsie were broader similar to Shuttle fault management activities except that, on shuttle, individual checklists are static paper entities. Compared to paper, the EPV contained multiple display features to assist with checklist navigation, such as highlighting the current "focus" line with a blue bar, coding already completed procedures by reducing their brightness and positioning them above the current focus line, and automatically skipping lines belonging to branches rendered irrelevant by conditional logic (e.g., if… then… else statements).

2.3. Besi

Figure 2-10 shows the fault management interfaces for Besi. Besi contains five areas: the Main Area, the Checklist Area, the System Status Area, the Root Cause Area, and the Fault Message Area. In Figure 2-10, the Main Area is displaying Besi's version of the EPS display. Although the performance impacts are not the focus of this report, in Besi, mode reconfiguration capabilities were embedded in the EPS Sum display itself. When the checklist called for a switch throw, the participant moved the cursor directly to the switch symbol within the line-diagram of the EPS display and toggled the switch state. Since there was no separate virtual switch panel that had to share the same display real estate as the system summary displays (as in Elsie), in Besi, the impacts of a switch throw to the system was immediately visible to the operator.

Besi was designed with three principles in mind:

1. Provide an interface with a model-based reasoner called the Hybrid Diagnostic Engine (HyDE), an ADAPT inference engine that diagnosed the parent of a C&W system event.

2. Provide more automated assistance with information navigation and information retrieval.

3. Lessen the information-processing load on the operator by automatically suppressing details deemed less relevant to the current operation. The suppressed details were still accessible to the operator by extra commands. For example, Besi's EPS Summary display (Figure 2-10) utilized more graphical representations than Elsie's. The default EPS display in Besi provided graphical representations for voltage (V), current (A), and battery temperature (T) for all active elements (see Figure 2-10). The height of the fills indicated the sensor readings, with C&W limits marked by red ticks, but no numerical data were presented by default. If a participant wished to view numerical values, s/he could bring them up by selecting the *View* edge key at the upper-left corner and then selecting the *#s: all* option in a pull-down menu. Then, the numbers appeared under the corresponding graphics. Also, Besi's EPS display automatically suppressed lines connecting inactive components to reduce display clutter. A participant could see the suppressed lines by, again, selecting the *View* edge key, and then selecting the *conn: all* option from the menu.

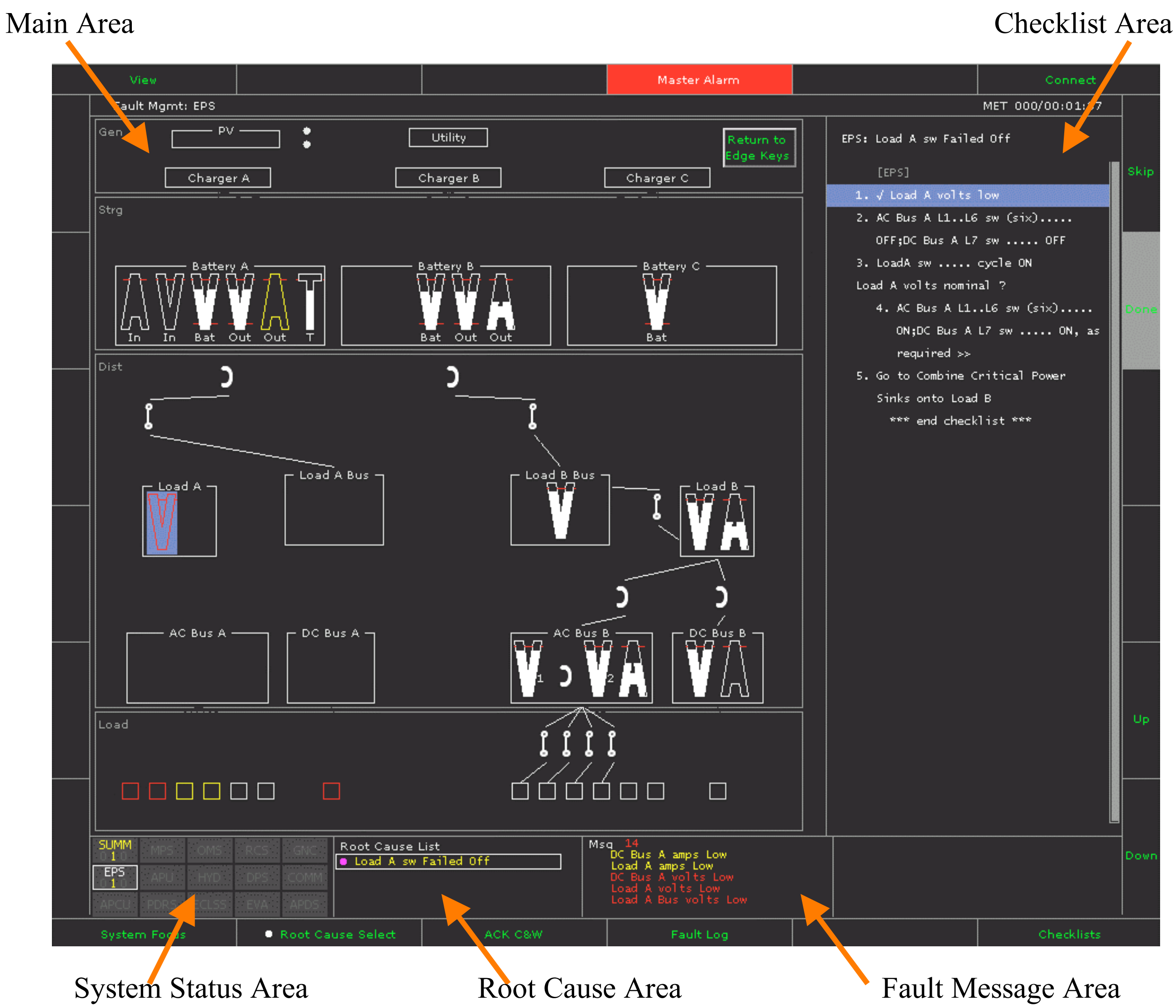


Figure 2-10. Besi.

Below Besi's Main Area were, from left to right, the System Status Area, the Root Cause Area, and the Fault Message Area. The matrix in the System Status Area remained dark gray if no out-of-limit sensor readings were detected. When any abnormal readings were sensed, the name of the system containing a problem showed up in yellow (caution) or red (warning) and the number of caution and warning messages was displayed. Thus, the System Status Area provided an at-a-glance indication of overall systems health. Also, moving the cursor to the SUMM, EPS, or ECLSS cell in the System Status matrix and selecting it brought up the Fault Sum, EPS, or ECLSS display, respectively. The Fault Sum and ECLSS display designs were common in Elsie and Besi (Figures 2-6 and 2-8). The Root Cause Area presented the proximal cause fault message as determined by HyDE. The adjacent Fault Message Area functioned in the same way as it did in Elsie. The Fault Log displays were also accessed via the edge key just like in Elsie. The electronic checklists were viewed in the Checklist Area on the right side of the Main Area. The Master Alarm edge key was at the top of the display.

A typical fault management process in Besi was as follows. As in Elsie trials, operators started out monitoring the Fault Sum and PFD displays. As soon as a malfunction was inserted, color changes for out-of-limits parameters appeared on Fault Sum (yellow for caution, red for warning) and fault messages appeared in the Fault Message Area; the Master Alarm was not yet issued. It usually took five to seven seconds for HyDE to complete its root-cause diagnosis. During this period, it was recommended, though optional, that operators check the Fault Sum display, and then either the EPS or the ECLSS display as appropriate, to examine which specific component(s) and sensor values were off nominal. Once HyDE completed the diagnosis, the fault message corresponding most closely to the proximal cause appeared in the Root Cause Area, and the Master Alarm was issued to draw the operators' attention.

In Besi, operators did not have to manually navigate to the target checklist by navigating through EPV menus. If they agreed with HyDE's proximal cause diagnosis, they would simply navigate to and click on Besi's *Root Cause Select* edge key label. This operation transferred cursor focus to the message that corresponded to the proximal cause fault message in the Root Cause Area. In turn, clicking on (selecting) this message immediately brought up the corresponding checklist in the EPV. Subsequent navigation through the checklist was done in an analogous way to Elsie, using the edge key labels on the right-hand side of the fault management display.

The content of the checklists for the same proximal cause were often different for Elsie and Besi, because HyDE automated many verification steps. Thus, instead of the lengthy diagnosing steps typically seen in the Elsie checklists, the Besi checklists often contained only a single verification step (e.g., "√ Load A volts low") to make sure that HyDE's computation was consistent with the operator's evaluation of the situation.

2.4. PFD Monitoring Task

A PFD monitoring task was also included to assess operators' ability to divide their attention between fault management activities and the more general flight monitoring activities that go on during ascent. The PFD, displayed in the top half of the 20-inch (portrait-orientation) monitor, included five flight parameters: altitude, velocity, G-meter, vehicle position, and thrust (see Figure 2-11). Each trial started with a simulated liftoff of the vehicle. Every 20 seconds (on average) the interior of one of the white background areas that housed critical flight parameters on the PFD changed to yellow. If the operator took no action, then after five seconds elapsed, the area changed from yellow to red. The area remained red for an additional five seconds and then, if the operator had still taken no action, returned to white. If, instead, the operator noticed the color change, he or she was trained to reach up and touch the indicator position directly while calling out its name (for example, if the G-meter box changed color, they would call out "G-meter"). Touching the colored parameter returned its area color to white immediately.

The operators were instructed to give equal importance to the PFD monitoring task and the fault management task, i.e. not to ignore the one in favor of the other. Nevertheless, humans are notorious for their limited ability to divide their attention between multiple simultaneous tasks. Our primary interest was in how the demands of the fault management task would impact PFD monitoring task performance. To the extent that the fault management (lower) display captured

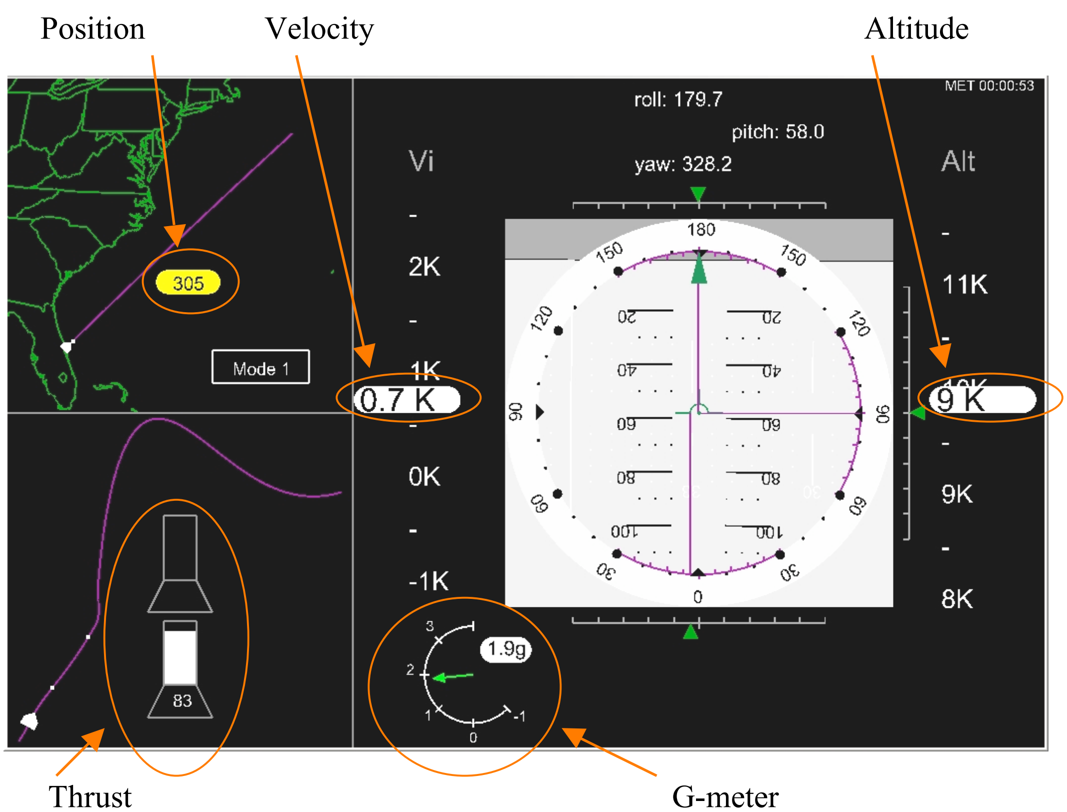


Figure 2-11. PFD

visual attention, for example, we would expect observes to be either slower to notice and respond to a color change on the PFD, or to miss color changes altogether. In fact, as described in the earlier report, operators did fail to notice significantly more color changes when working malfunctions with the less automated Elsie than with Besi. This result was supported by a preliminary analysis, which revealed that operators spent a greater proportion of time looking at the PFD while working malfunctions with Elsie than with Besi.

Clearly, Besi reduced visual tunneling on the fault management displays. However, just how our operators implemented this reduction could not be ascertained from the rather coarse analyses in the earlier report. In any multi-tasking environment, where the information associated with the constituent tasks is located in discrete visual locations, operators must develop and implement a strategy for how to distribute and coordinate their information acquisition activities across time. How often does one task get interrupted to service the other? What determines when an interrupt decision was made? Does it depend on the particulars of what is being processed “at the moment”, or does it reflect a top-down strategy that was invariant with respect to moment-by moment activities and demands? The new and much more fine-grained analyses of eye movement behavior reported here allow us to begin to address these issues.

2.4.1. Operators

Eight operators, all instrument-rated pilots, were recruited for the study. The operators included seven males and one female, and their ages ranged from 24 to 54 (average of 37.5). Their total flight times ranged from 230 to 21000 hours, and the instrument-flight times ranged from 68 to 2000 hours. All operators were right-handed and had normal or corrected vision (i.e., 20/40 or better). Each operator received approximately twelve hours of training, including four hours of reading assignments, four hours of classroom lecture, and four hours of hands-on practice working EPS and ECLSS systems malfunctions using a laptop-computer-based trainer. Each operator was required to pass a final exam immediately prior to his or her first data-collection session.

2.4.2. Data Collection

Data collection was split into two sessions. Each session consisted of seven simulated ascents with one display format suite (either Elsie or Besi). Half of the operators completed seven Elsie trials in the first session followed by seven Besi trials in the second session. For the other half, the session order was reversed. Scenario orders were counterbalanced across participants.

Table 2-1 lists the fourteen malfunction scenarios used. Scenario pairs #5 and #6, #7 and #8, #9 and #10, and #13 and #14 were symmetric, and used in the different ACAWS displays within a participant. For instance, a participant who was assigned scenario #5 for Elsie was assigned scenario #6 for Besi, or vice versa. Scenarios #11 and #12 were identical. Scenarios #3, #4, and #11 through #14 contained multiple malfunctions. In these scenarios, the second malfunction occurred 55 to 90 seconds after the occurrence of the first malfunction so that the participant was still working on the first malfunction when the second one occurred.

Scenarios #1 through #4 formed one group that provided 2 x 2 conditions, real failure vs. sensor failure (see Hayashi et al., 2007 for details) and single malfunction (lower workload) vs. multiple malfunctions (higher workload). The switch sensor failures in #2 and #3 caused a false alarm when the switch was actually functioning correctly. For the Besi trials, the sensor failure also caused HyDE to misdiagnose and generate a root cause that did not exist (the original HyDE algorithms were able to distinguish sensor failures from actual component failures, but for this study, HyDE was intentionally modified to generate a false root cause in these particular cases so that the participants' responses to an inappropriate diagnosis could be studied.) Due to the different experiment designs, scenarios #1 through #4 were analyzed separately from the rest of the scenarios.

Each trial started from liftoff and ended at eight, ten, or twelve minutes into flight, whichever cut off came first after the operator completed the malfunction resolution. If the operator was unable to resolve the malfunction, the trial was ended at twelve minutes. During the trials, operators' ACAWS display commands (edge key navigations, checklist navigations, switch throws, etc.) were recorded and time stamped. Likewise, operators' touches of the PFD were recorded. All trials were video recorded.

Table2-1. Malfunction Scenarios

Scenario #	Malfunction(s)
1	A/L1 sw mismatch
2	B/L1 sw mismatch (false alarm)
3*	1) Load B sw mismatch (restorable) 2) A/L2 sw mismatch (false alarm)
4*	1) Load A sw mismatch (restorable) 2) B/L2 sw mismatch
5	DistAA sw mismatch (restorable)
6	DistBB sw mismatch (restorable)
7	Battery A volts low
8	Battery B volts low
9	Inverter A failure
10	Inverter B failure
11*	1) Inverter A failure 2) Battery A volts low
12*	Same as 11
13*	1) Battery A volts low 2) Battery B volts low
14*	1) Battery B volts low 2) Battery A volts low

* : Multiple-malfunction scenarios

Immediately after each trial, questionnaires for TLX (Hart & Staveland, 1988) and modified Bedford (Roscoe, 1984; Huntley, 1993) workload scales were administered via a computer interface. Following the 7th and 14th trials (i.e., at the end of each session), operators provided their TLX pairwise comparisons via an electronic questionnaire. After completing both sessions, they provided written answers to questions regarding display usage, display preference, and user interface design.

2.5. Summary of Previously Reported Findings

The major findings from the previous report are as follows.

- Table 2-2 summarizes the performance accuracy grading results for Elsie and Besi for each scenario. The trials were graded *Correct* if all proper checklists were used AND all switch throws were performed correctly. Trials where some or all of the proper checklists were not used or the final switch configuration was incorrect were graded *Failed*. All other trials (i.e., all proper checklists used and final switch configurations correct, but some improper switch throw(s) occurred during the process) were graded *Good*. Besi trials resulted in more *Correct* and *Good* performances combined (i.e., *non-Failed* performances) and less *Failed* performances than Elsie trials (both trends were statistically significant).

Table 2-2. Malfunction Management Procedure Accuracy by Scenario

Scenario #	Malfunction(s)	Elsie			Besi		
		Correct	Good	Failed	Correct	Good	Failed
1	A/L1 sw mismatch	2	1	1	3	1	0
2	B/L1 sw mismatch (sensor failure)	3	0	1	3	1	0
3*	1) Load B sw mismatch (restorable) 2) A/L2 sw mismatch (sensor failure)	1	0	3	3	0	1
4*	1) Load A sw mismatch (restorable) 2) B/L2 sw mismatch	3	0	1	4	0	0
5	DistAA sw mismatch (restorable)	4	0	0	4	0	0
6	DistBB sw mismatch (restorable)	4	0	0	4	0	0
7	Battery A volts low	3	0	1	3	0	1
8	Battery B volts low	4	0	0	3	0	1
9	Inverter A failure	4	0	0	3	0	1
10	Inverter B failure	4	0	0	4	0	0
11*	1) Inverter A failure 2) Battery A volts low	2	0	2	2	2	0
12*	Same as 11	0	2	2	1	0	3
13*	1) Battery A volts low 2) Battery B volts low	2	0	2	4	0	0
14*	1) Battery B volts low 2) Battery A volts low	0	2	2	2	2	0
	Total	36	5	15	43	6	7

* : Multiple-malfunction scenarios

- The malfunction resolution times (RT) for the single-malfunction scenarios (#5 through #10) showed that Besi significantly shortened the RTs for scenarios #5/#6 and #9/#10. Further analyses on different phases within the RTs indicated that Besi significantly shortened the time spent for initial diagnosis.

- Analyses of the PFD color-change monitoring task during the on-task time (the time during which the participant was working on a malfunction) indicated that Elsie trials tended to result in more task misses than Besi trials. The results provide preliminary evidence that attentional tunneling on the malfunction procedures was more severe in Elsie trials than Besi trials.

- Preliminary eye-movement analysis based on a coarse parsing of the display into two regions – the upper PFD and the lower fault management area - revealed that the operators monitored the PFD for 23.9% of the total on-task time in Elsie trials, versus 30.5% in Besi trials. The difference was statistically significant.

- Both NASA TLX workload scores and the modified Bedford workload scores for the paired scenarios (#5 through #14) showed that participants experienced significantly less workload with Besi than with Elsie.

- Operators preferred Besi significantly more than Elsie for the diagnostic process and the recovery process. They also preferred Besi's switch symbol representation significantly over Elsie's virtual switch panel representation. They marginally preferred Besi's graphical representation of parameter values over Elsie's textual parameter representation.

2.6. Focus of the Present Report

Many issues concerning display usage and information acquisition strategies cannot be answered with these conventional human factors performance measures. In this report, we provide more fine-grained analyses of oculomotor behavior that enable a more direct assessment of operators' information acquisition strategies and display usage patterns. The results provide clearer guidance for the design of fault management displays and a better understanding of the impacts of the various forms of Elsie automation.

While some of the new analyses encompassed all phases of the fault management task, our primary focus is on the diagnostic and checklist selection phases, as these provided the most direct assessment of the impact of automating the proximal-cause diagnosis and checklist retrieval operations. To provide the appropriate context for these analyses, we will first make more explicit some of the links between the information display requirements of our multi-tasking environment and our display layout and display navigation schemes.

2.6.1. Display Real Estate and Display Availability

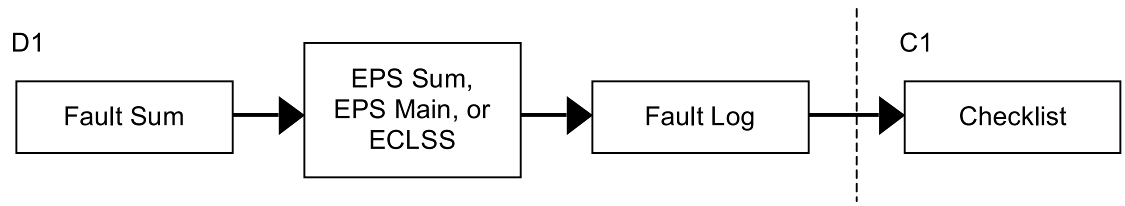
Prior to the insertion of a malfunction, the operator's tasks were to monitor vehicle flight status information, detect and respond to color changes on PFD flight parameters, and monitor systems operations for any operational anomalies. The information display requirements for this mix of tasks were fully satisfied by the PFD and Fault Sum displays. As soon as a malfunction occurred, the multi-tasking demands on the operator increased, as did the requirements for information display. For example, in Elsie, operators were required to make a proximal-cause diagnosis of the off-nominal event, and were trained to utilize information from several sources to support their proximal-cause fault diagnosis. Due to display real estate limitations, some of these sources could be viewed simultaneously, others only sequentially. Figure 2-12 illustrates the suggested time course of display usage in Elsie and Besi, as per the training regimen, for display formats that could only be viewed sequentially. In both operational concepts, operators were trained to first utilize the Fault Sum display, which always occupied the main display area at the time the malfunction(s) occurred. As shown in Figure 2-12, operators were trained to acquire further insight from the EPS graphical displays (EPS Sum or EPS Main in Elsie; EPS in Besi).

At this point, the information requirements for Elsie and Besi started to deviate. In Elsie, the proximal-cause diagnosis amounted to selecting a C&W fault message, as these messages did “double duty” as labels for the off-nominal EPV checklists. Thus, operators generally had to replace the EPS displays in the Main Display Region with the C&W Fault Log page (only two scenarios contained the parent fault message in the list of messages in the fault message region). By contrast, Besi provided a candidate proximal-cause fault message, and the act of selecting that message brought up the corresponding EPV checklist automatically. As a result, bringing up the Fault Log page was virtually unnecessary in Elsie.

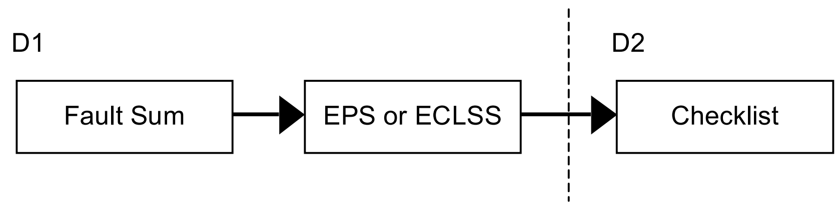
One source of fault-management information, the subset of fault messages presented in the lower message area, was always in view. Thus, there were always at least three sources of task-relevant information available for viewing at any one time: the PFD (for the color-change task), and for the fault-management task, any one of a number of graphical fault-management related displays in the main display area, and the fault messages in the lower C&W message area. If we also include the edge key labels used for display navigation and display selection, there were actually four distinct regions competing for operator’s attention. How operators coordinated and scheduled information acquisition activities between these regions is a major focus of this report. The results are relevant to several fundamental issues relating to fault management operational concept definition and interface design.

2.6.2. Display Usage: Where did the time go?

Although the architecture of our display navigation scheme forced strictly sequential processing of Fault Sum, EPS summary displays, and Fault Log, the precise amount of time spent looking at these displays was entirely at the operators’ discretion. With multiple areas simultaneously competing for attention, an objective measure, such as the time stamps of the edge key presses that swapped out these displays, does not provide a definite measure of actual time-on-display. Only analyses of eye fixations can determine how much use operators made of these multiple sources of information, most notably the textual information provided by the fault messages, compared to the more graphical information provided by the Fault Sum, EPS Sum (Elsie) or EPS (Besi) displays. Of the total time taken to diagnose the malfunction, what proportion was devoted to each information source?



(a) Elsie



(b) Besi

Figure 2-12. Display Usage Flow during Initial Diagnosis

2.6.3. Information Acquisition Strategies and Display Real Estate Requirements

In next generation spacecraft, checklists of off-nominal procedures will migrate from paper to an EPV. In addition, since there will be little or no display of C&W system information on dedicated hardware panels (such as panel F7 on the shuttle); virtually all C&W information will be depicted electronically. The net increase in demand for electronic real estate to support real-time fault management operations will have to be accommodated within a cockpit with much less display real estate available than the shuttle. That supply will be further restricted during the dynamic flight phases, when the crew will need continuous access to GNC and safety-critical systems summary displays (such as the main propulsion system summary display). Thus, the identification of the minimal display real estate requirements necessary to support fast and accurate fault management operations is a critical issue.

When a display environment presents the operator with multiple distinct sources of task-relevant information, as was the case with our operational concepts, the operator has considerable leeway as to how he or she organizes, coordinates, and sequences individual information acquisition events (IAEs) from each source. Analyses of these strategies offers preliminary guidance for how much (or what forms of) fault management information should be provided simultaneously, and what forms might safely share a single region of display real estate (and therefore not available for simultaneous viewing), without compromising task performance.

As we noted, in our display configurations, there were at least two sources of fault-related information available at all times. Importantly, the information in these sources remained the same throughout the diagnosis phase of the fault management process (D1). Since the information content was static, two information acquisition strategies were possible (Droll, Hayhoe, Treisch, & Sullivan, 2005; Droll and Hayhoe, 2007). One strategy would be to select (fixate on) a display region and fully encode all the task-relevant information from that information source into working memory in the single IAE. This would be revealed by the existence of an initial IAE to each relevant region, and few or no follow-ups.

On the one hand, this "all-in-one" encoding strategy would place high demands on working memory resources; for example, if the diagnostic process was supported by cross-checking information from the multiple displays with relevant information (such as, for example, the EPS system summary displays and the fault messages in the fault message area), this cross checking would entail comparing information extracted from the current IAE (on the currently attended ROI) with information stored in working memory from the previously attended areas. On the other hand, "all-in-one" encoding would minimize eye movements and the accompanying need for scheduling and executing saccades, behaviors that have been inferred to contribute to operator workload (Droll et al., 2005). In addition, comparing information extracted from a current IAE with information from memory might be more efficient than cross-checking with a visual source, as that information has to be re-attended and re-encoded. These steps are obviously eliminated when the information is available in memory.

The alternative strategy would be to minimize the demands on working memory by acquiring information from available sources in a more incremental fashion, across several distinct IAEs, separated by IAEs to other regions of interest. For example, following Droll and Hayhoe (2007), operators might choose to relieve the working memory load by encoding only a subset of the total information available in a particular region of interest on any particular fixation. They might then have to cross-check new information, acquired during the current IAE, against information in the other region that was not relevant when they visited it earlier, and had therefore not been encoded. This strategy would be revealed by the presence of multiple IAE's to the same "region of interest" (ROI) (of course, this strategy might simply be mandated by limitations in human working memory resources, or if previously encoded information is subject to short-term memory decay).

Whether operators do or do not return to ROIs already visited has direct implications for fault management-related requirements for display real estate. If operators spontaneously adopt the "full encoding" strategy, and revisit only rarely, the efficiency with which they perform the fault management task should not be greatly impacted if relevant information sources have to time-share the same section of display real estate, reducing the overall display real estate requirements of the fault management task. On the other hand, if revisiting is common, the implication would be that an environment in which multiple sources of fault management information were available for viewing simultaneously would better support fault management. This would, of course, raise the display real estate requirements needed to optimize task performance.

2.6.4. Multi-Tasking Strategies

In our earlier report, we noted that operators missed significantly fewer color changes on their PFD symbology while working malfunctions with Besi than with Elsie, indicating that Besi influenced the divided attention strategy between the fault management task and the PFD-based task. Analyses of eye movements can allow us to distinguish between three more specific hypotheses. The performance benefit could be due to 1) increased duration of individual PFD fixations, 2) increased number of IAEs to the PFD per unit time (higher sampling frequency) or 3) both higher sampling frequency and longer durations.

3. Eye-Movement Data Processing

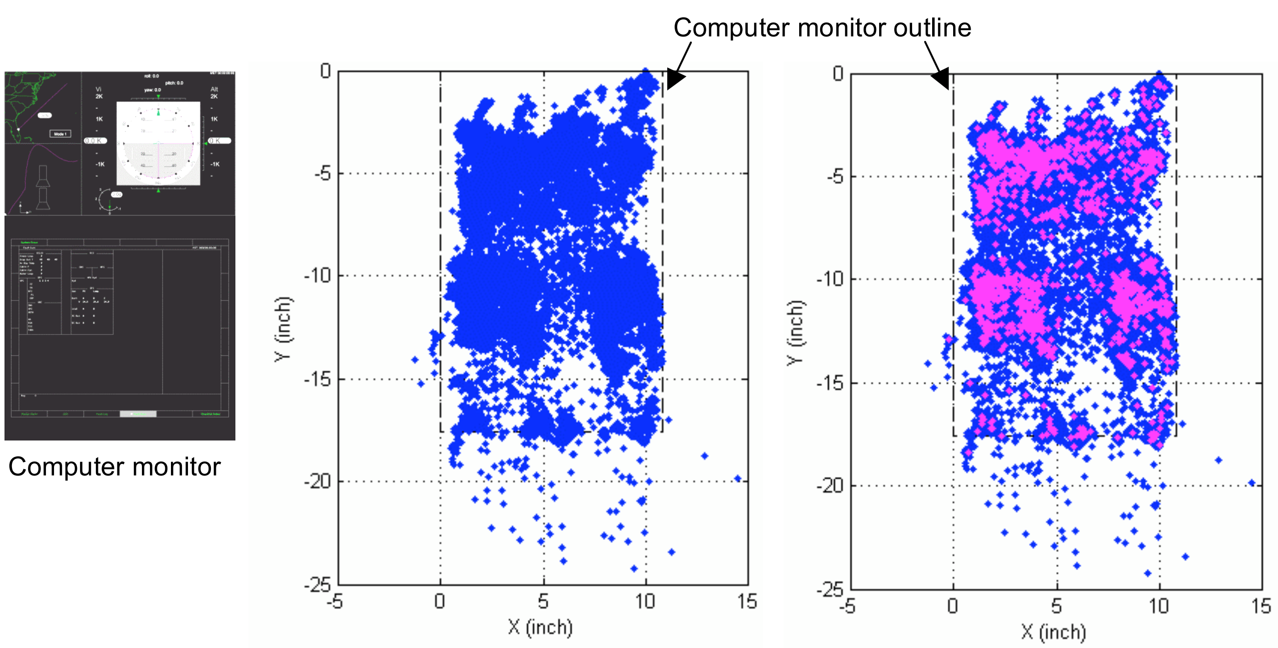
3.1. Raw Eye-Movement Data

The ISCAN eye-and-head-tracking system measured the head positions (x, y, and z), head rotations (azimuth, elevation, and roll), pupil center locations (x and y), and corneal reflection locations (x and y) at the sampling rate of approximately 60 Hz. These measurements were used to compute the participant's line-of-sight eye angles. The 3D coordinates of the four corners of the computer monitor had been registered in the eye-and-head-tracking system, and with that information, the ISCAN's software computed the Plane Intersection Coordinates (PICs), that were the x-y intersection coordinates on the plane (i.e., the computer monitor surface) at which the line of sight penetrates the plane. The PICs were computed for each sampling (i.e., approx 60 Hz). These PICs were the starting point of the following eye-movement data analyses, and are referred to as raw eye-movement data in this report.

Once before and once after each trial, the participants were asked to look at each of the three-by-six grid reference points for about two seconds each. This process is called external calibration. The external calibration data were recorded and used for determining the borders between regions of interest (see Section 3.3).

3.2. Identifying Fixations

The raw eye-movement data included fixations and saccades (rapid eye movements in between



(a) Raw eye-movement data (blue)

(b) Fixation centroids (magenta) superimposed on raw data (blue)

Figure 3-1. Identifying Fixations

fixations). For the purpose investigating participants' display usage, first, only the fixation data points were extracted. The fixation criteria used in this study were as follows: For successive data points to be classified as part of the same fixation, they had to be within 0.76 inch both horizontally and vertically from the centroid of the fixation, and the duration of the data point had to be 100 msec or longer. The algorithm ignored noises of up to five consecutive samplings. Any data points that did not satisfy the fixation criteria were considered saccades and removed from the analysis. Figure 3-1 shows examples of the raw eye-movement data (blue) collected during a single trial, and their fixation centroids (magenta) superimposed on the raw data.

3.3. Classifying into Regions of Interest

Next, each fixation was classified into one Region of Interest (ROI). The area of the computer monitor was divided into four primary ROIs, which are PFD, Main, Checklist, and Message (Figure 3-2). All borders were either horizontal or vertical (i.e., no tilt). The initial positions of these borders were selected based on the external calibration data, and then the border positions were fine-tuned around the initial positions so that the numbers of data that the borders crossed were minimized. Figure 3-2 shows an example ROI classification of the fixations.

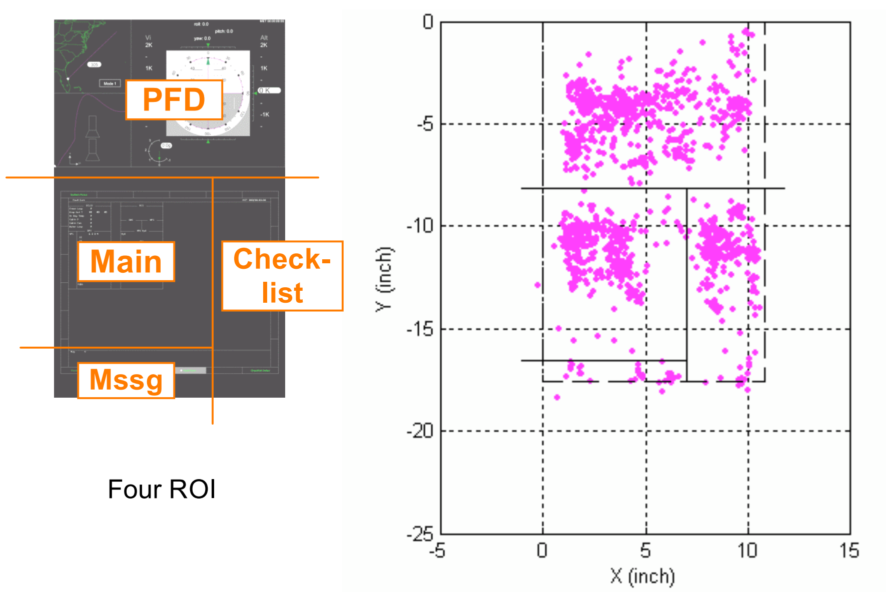


Figure 3-2. Classifying into ROI

Note that the ROI corresponding to the main display area contained different displays at different times. Thus, the Main ROI was further classified into individual displays of Elsie or Besi. Table 3-1 lists the ROIs for Elsie and Besi.

Also note that the Message ROI in the Elsie trials contained the Fault Message Area and the edge keys located at the bottom of the screen. In the Besi trials, the Message ROI included three areas at the bottom, which were System Status, Root Cause, and Fault Message Areas, and the edge keys at the bottom.

Table 3-1. ROI

Elsie ROI	Besi ROI
▪ PFD ▪ Fault Sum* ▪ EPS Displays (combining EPS Sum, EPS Main, and EPS Loads displays)* ▪ EPS Switch Panels (combining EPS Dist Switch Panel and EPS Loads Switch Panel)* ▪ ECLSS* ▪ Fault Log (combining p.1 through p.3)* ▪ Checklist ▪ Message	▪ PFD ▪ Fault Sum* ▪ EPS* ▪ ECLSS* ▪ Fault Log (p.1 - p.3)* ▪ Checklist ▪ Message

* : Main ROI group

Finally, all temporally adjacent fixations within the same ROI were grouped together to form a single IAE. In other words, each IAE represents one distinct visit to one ROI, and usually contained multiple fixations. The start time and duration of each IAE were computed, and used for the analyses.

3.4. Definitions of RT Phases and On-Task Times

Since the study's major interest was to understand operators' display usage during the fault management process, most of the eye-movement data analyses were performed only on the data during the time the participants were working on a malfunction or malfunctions.

The time from the appearance of the first fault message to the completion of the procedures checklist is called malfunction resolution time (RT). Strictly speaking, RT was the sum of the durations of five quasi-distinct phases, namely, Diagnosis 1 (D1), Checklist 1 (C1), Diagnosis 2 (D2), Checklist 2 (C2), and Recovery (R), each of which was associated with tasks of different natures. Most of the analyses were performed separately for each RT phase. Table 3-2 describes the five phases.

Computing and comparing the RT phase durations became difficult when the participants deviated from standard procedures (the trials graded *Good* or *Failed*) or interleaved two different procedures (multiple-malfunction scenarios). For those reasons, the durations of the RT phases were calculated for only the single-malfunction trials (scenarios #5 through #10) graded *Correct* most of the times. The only exceptions were the analyses on the data during D1 (Sections 4.1.1 and 4.3). The reason why all the data, that were the scenarios #5 through #14 trials regardless of the grade, were included for only the D1 phase was that, unlike during the other phases, defining errors was not as relevant to the D1 phase. Malfunction performance grading was basically determined by the switch-throw and checklist-navigation performances; neither of which was a part of D1. Among the five phases, D1 was the only phase where the participants used the fault management displays in a quasi open-loop fashion. Therefore, the D1 phase was particularly interesting for this study's purposes, and the exception helped maximize the amount of usable data.

Table 3-2. Five phases of RT

Phases	Descriptions	From	To
Diagnosis 1 (D1)	The operator assessed the situation and made a root cause determination before selecting the checklist to go to.	The first fault message was displayed.	*Checklist Index* edge or *Root Cause Select* edge key (Besi) was selected.
Checklist 1 (C1)	The operator navigated to the proper checklist. This process was manual in Elsie, and automated in Besi.	*Checklist Index* edge or *Root Cause Select* edge key (Besi) was selected.	The correct checklist was displayed.
Diagnosis 2 (D2)	The operator followed the checklist instructions to perform the diagnosis steps. Besi's checklists usually contained fewer diagnosis steps than Elsie's.	The correct checklist was displayed.	Completion of the last line of the diagnosis steps was acknowledged, or the line that instructs transitioning to the recovery checklist (if there is a separate recovery checklist) was acknowledged.
Checklist 2 (C2)	The operator navigated to the recovery checklist. If there was no separate recovery checklist, C2 duration was zero. Again, going to the recovery checklist was manual in Elsie and automated in Besi.	The line that instructs transitioning to the recovery checklist was acknowledged.	The correct recovery checklist was displayed.
Recovery (R)	The operator performed the instructed switch throws to reconfigure the system for recovery.	Completion of the last line of the diagnosis steps was acknowledged, or the correct recovery checklist was displayed.	Completion of the last line of the recovery steps was acknowledged.

The other concept related to the RT is *on-task time*. The time between the start of D1 and the end of R (if it was a multiple-malfunction scenario, the R of the second malfunction) was simply the time where the participant was working on some aspect of the malfunction management (no matter what it was). This period will be called *on-task time* in the following analyses. The on-task time never had a break, even for the multiple-malfunction trials. The advantage of the on-task time over the RT phases was that the on-task times could be computed easily even for the multiple-malfunction trials and the trials graded *Good* or *Failed* (as long as there is some marker that indicates the end of the on-task time). The definitions of the on-task time were somewhat looser and more relaxed than those of the RT phases, but may be sufficient for certain type of the analyses, especially when the focus was not the usage of the ACAWS displays themselves. For instance, most analyses of the PFD task performance (the task concurrent with the ACAWS malfunction-management tasks) used the on-task time rather than the RT phases.

4. Results

4.1. Percent Display Usages

4.1.1. Aggregate Display Usage during D1

Figure 4-1 shows the percent usages of the displays participants examined during the D1 phase of the paired trials (scenarios #5 through #14). All paired-trial data, regardless of the number of malfunctions and their grading, were used for the D1 phase. Note that we used different criteria for data inclusion from the rest of the RT phases, where data from the only single-malfunction trials graded *Correct* was used (See Section 4.4 for the rationale). The total length of each bar was adjusted to scale with their grand average durations (i.e., 43 sec for Elsie, 24 sec for Besi).

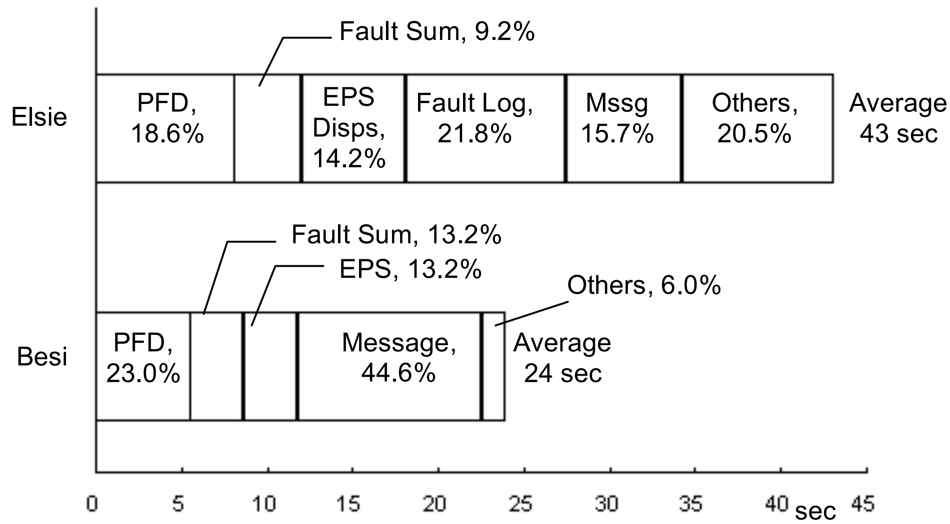


Figure 4-1. Display page usage during D1

The EPS Displays ROI in Elsie included the EPS Sum (11.5%) and the EPS Main (2.4%). No participant used the EPS Loads during D1 of the Elsie trials. No participant used the ECLSS display during D1 of these scenarios. The *Others* category in Elsie included dwells in the Checklist ROI (6.2%), the system list shown in the Main Area after selecting the *System Focus* edge key (5.9%), and a blank screen displayed in the Main Area after selecting EPS in the system list (3.4%). The "*Others*" category in Besi included dwells in Fault Log (4.3%) and the Checklist ROI (0.5%). The Checklist ROI were grouped in the *Others* category for D1, because this ROI remained mostly blank during D1, except for the Mission Elapsed Time (MET) indication in the upper-right corner and the *Checklist Index* edge key in the lower-right corner. The remaining percentages of the "*Others*" category were composed of saccades among the ROI, or noise.

D1 is the RT phase where the participants diagnosed the malfunction, based partly on instructions, but largely using strategies of their own choosing. Therefore, their display usage during D1 was particularly interesting. The Elsie graph reveals that over one third of diagnosis time was spent viewing either Fault Log or the Fault Message Area, the two areas that provided text-based information about the malfunction (note, however, that the Fault Message Area also included the edge keys; this issue will be revisited in Section 4.3). Similarly, in Besi, the text-based ROIs (System Status, Root Cause Box, and Fault Message Area) accounted for over 40% of the total D1 time. Note that this one ROI contained not only Besi's proximal cause message in the ACAWS dialog box, and raw (unfiltered) fault messages in the Fault message area, but also the C& matrix and the edge keys used to navigate among displays in the Main Display Area. The multiple sources of information may explain why this area received so many fixations during D1 in Besi.

4.1.2. Percent Display Usages during C1

Figure 4-2 plots the percentage of total viewing time to the Elsie displays during phase C1 of the single-malfunction trials (scenarios #5 through #10). Unlike in Figure 4-1, only data from the trials graded *Correct* were included. The number of qualified trials was 23 for Elsie and 21 for Besi. In Besi trials, operator selection of the proximal cause fault message brought up by the associated checklist in the EPV automatically, fixing the duration at 0.5 sec. Because they were such short durations, the display usages were not examined for C1 for Besi (however, the total length of the Besi bar was shown in the chart for comparison). In sharp contrast, manually navigating through the EPV menu to bring up the targeted checklist in Elsie took an average of 20 sec.

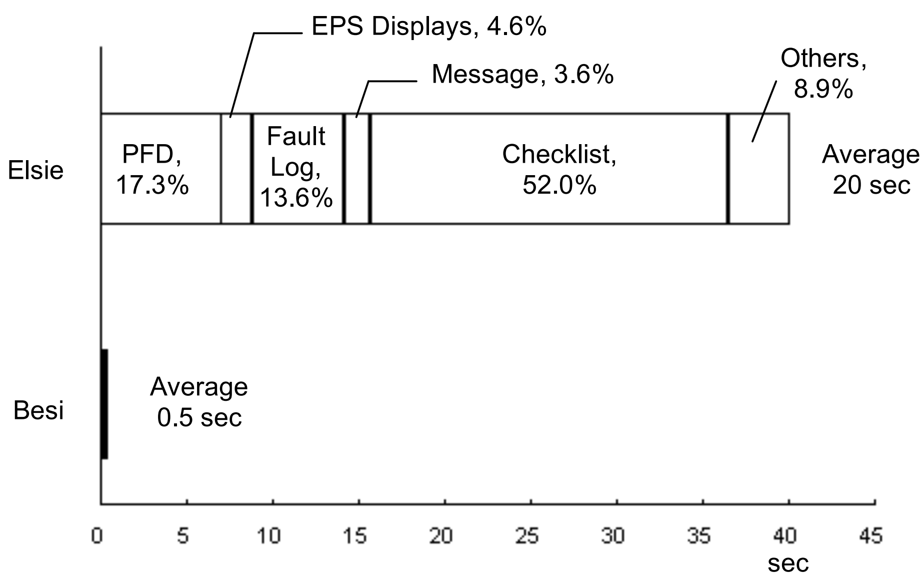


Figure 4-2. Display page usage during C1

The EPS Displays ROI was comprised of the EPS Sum (0.3%), the EPS Main (2.1%), and the EPS Loads (2.1%). The *Others* category included Fault Sum (1.2%), the EPS Switch Panels (2.4%) and lesser amounts due to saccades or noise.

Importantly, the Elsie graph reveals that more than half of the total C1 time was spent looking at the checklist menu, which contained a list of checklist titles for EPS malfunctions. However, operators also continued to examine the list of fault messages on the Fault Log and the messages contained in the Fault Message Area, presumably to help select the proper checklist name in the checklist menu.

Display usage for the latter stages of the fault management task are included in Appendix A for those who what to compare with these early stages.

4.2. PFD Monitoring

Table 4-1 summarizes the participants' PFD look statistics separately for operations concept (Elsie versus Besi) and operator workload (single malfunction versus multiple malfunction runs) during the time that operators were actively involved in fault management activities. For this analysis, all trials, regardless of the scenarios or grading results, were included, as the focus here was on understanding how the operations concept (Elsie versus Besi) influenced the division of attention between fault management and the PFD-based color noticing task.

Table 4-1. PFD Look Statistics during On-Task Time

Scenario Difficulty x ACAWS	N	% time of PFD look	Number of PFD looks per minute	Average PFD look duration [sec]
Single-mal Elsie	24	20.6 %	14.6	0.86
Single-mal Besi	24	26.5 %	17.1	0.95
Multiple-mal Elsie	16	16.5 %	11.9	0.86
Multiple-mal Besi	16	20.8 %	14.1	0.90

As shown in the column labeled "Number of PFD looks per minute", during the period that operators were actively working a malfunction, they glanced at the PFD more often with Besi than with Elsie. Summing the amount of time that operators "borrowed" from fault management to look at the PFD, and expressing that summed value as a percentage of the total fault resolution time, the summed value (called "% time of PFD look" in the Table) was higher for Besi than for Elsie. Paired t-test results indicated significant effects of operations concept on both measures in the single-malfunction scenarios (scenarios #5 through #10; $t(23) = 3.36$, $p < 0.01$ for % PFD looking time; $t(23) = 3.96$, $p < 0.01$ for the number of looks per minute) and the multiple-malfunction scenarios (scenarios #11 through #14; $t(15) = 2.95$, $p = 0.01$ for the % time; $t(15) = 3.24$, $p < 0.01$ for the number of looks per minute). The average duration of the looks to the PFD did not show any statistically significant effect of operations concept.

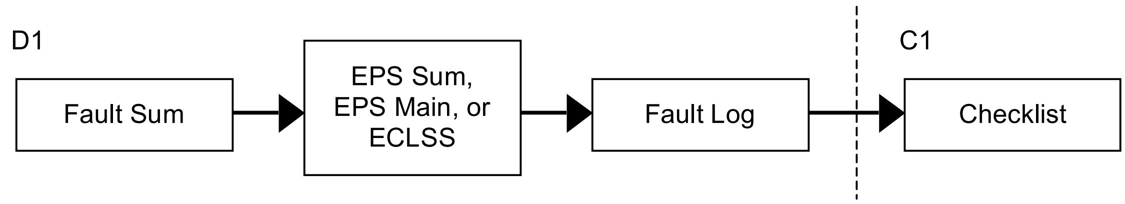
These analyses are broken out further in Appendix B for the latter phases of the fault management task, for those who want to check the consistency of PFD task-related behavior between all fault management phases.

4.3. Further Breakdown of Display Usage during D1

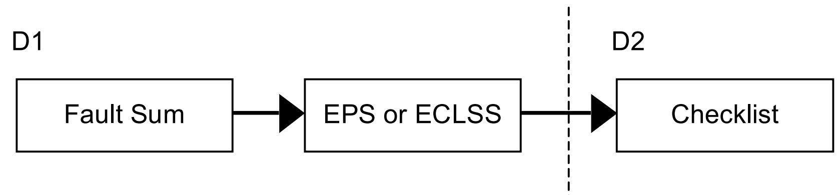
Section 4.1 provided an aggregate look at eye movement behavior during D1. However, that analysis finessed the fact that over the course of a typical diagnosis phase, three very different display formats appeared in the Main Display Area. We now “drill down” to a more fine-grained analysis of eye fixation behavior separately for each format.

4.3.1. Outlines of the Display Usage Flow during D1

As touched upon in Section 1.4, even though there was no rigid order of the display usages



(a) Elsie



(b) Besi

Figure 4-3. Outlines of Display Usage Flow during D1

imposed to the participants during the D1 phase, there was a suggested order that naturally made sense. Figure 4-3 illustrates the outlines of the suggested usage for each display format during D1.

For both Elsie and Besi, the flow naturally started with Fault Sum, which was always present when the malfunction occurred. Through color coding of off-nominal elements, Fault Sum offered “at a glance” information concerning which system, EPS or ECLSS, contained the malfunction, and which components of these two systems were affected. Although we included ECLSS (loads) malfunctions during our training phase, all malfunctions were in the EPS system during the data collection runs. Operators were instructed to replace Fault Sum with EPS summary displays, as this would give them further insight into EPS status and support their efforts to determine the fault message corresponding to the proximal cause.

At this point, the natural flow for Elsie and Besi deviated. For Elsie, operators were instructed to bring up the Fault Log display, again, to locate the fault message most closely associated with the proximal cause. By contrast, Besi enabled selection of the “most-proximal” fault message automatically (in practice, no operator brought up Fault Log when working malfunctions with Besi). Thus, in Figure 4-3, Fault Log was omitted from the D1 display usage flow for Besi.

Figure 4-4 plots the probabilities of the use of each display across D1 for Elsie. All data from scenarios #5 through #14, regardless of the grading results, were included. For the multiple-malfunction trials (#11 through #14), only the data from the first malfunction were included. The horizontal axes were normalized mission elapsed time (MET) where 0% indicates the beginning of D1, and 100% indicates the end. These plots reveal that, in compliance with

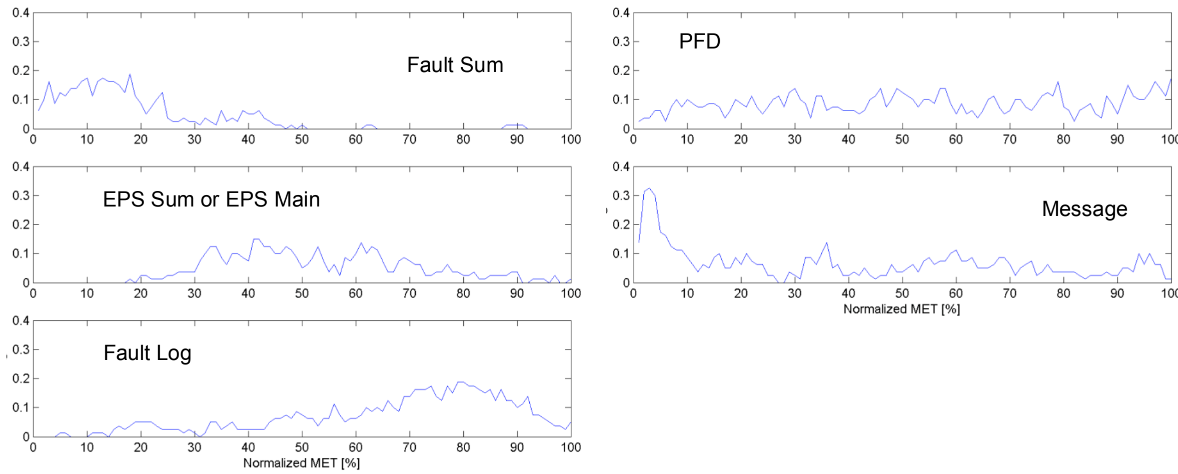


Figure 4-4. Elsie Display Usage Probabilities during D1

training, Fault Sum viewing dominated the beginning of the period, EPS Sum or EPS Main dominated the middle, and Fault Log dominated toward the end. Importantly, operators interrupted their fault management activities to examine the PFD at a fairly constant rate throughout D1. The Message area was also examined at a fairly uniform rate except for the sharp peak at the very beginning, when the operators were looking at and extinguishing the visual Master Alarm.

Figure 4-8 shows the same plots for Besi. Again, all the data from the scenarios #5 through #14 were included. For the multiple-malfunction trials (#11 through #14), only the data from the first malfunction were used. The plots show that the Fault Sum was used during the beginning of D1, while the EPS was consulted during the middle to the end of D1. The PFD and Message areas

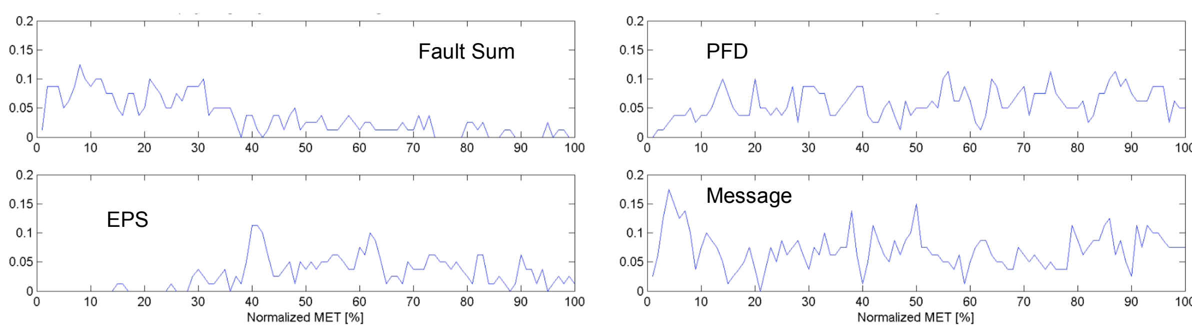


Figure 4-5. Besi Display Usage Probabilities during D1

were examined fairly uniformly throughout.

4.3.2. Display Usages and Fixation Transitions by D1 Sub-Phase

Figures 4-4 and 4-5 indicate that D1 could be divided roughly into distinct sub-phases of information acquisition activity, depending on which display occupied the Main Area: A Fault Sum sub-phase, an EPS display(s) sub-phase, and a Fault Log (Elsie) sub-phase. When Fault

Sum was displayed in the Main Area, the EPS-related and Fault Log displays were unavailable. Therefore, the probabilities of the usage of the displays not shown were zero. Furthermore, to "swap-out" display formats in the Main Area, operators needed to navigate to, and select, edge key labels. That means that the probabilities of direct visual transitions between Main-Area display formats were always zero. In this section, display usage during D1 was examined separately by sub-phases.

In addition, in the omnibus analyses reported in Section 4.1.1, there were ambiguities with respect to the Message Area, which included the edge keys as well as the Fault Message Area (and, in case of Besi, the Root-Cause and the System Status Areas). Viewing fault messages and viewing edge keys supported two very distinct tasks. To differentiate between looks to the edge keys and fault messages, therefore, video recordings of the operators during the trials were consulted. During periods when it was clear that operators were moving the current cursor focus among the edge key labels, the corresponding fixations were classified as edge key fixations. The remaining fixations in the Message Area, including ambiguous cases, were classified as fault message area fixation. Since video examination was a labor-intensive process, only D1 fixations were classified with this method.

The Fault Sum Sub-Phase

Figure 4-6 plots the participants' display usages during D1 when Fault Sum was displayed in the Main Area. The data from all the paired trials (i.e., scenarios #5 through #14), regardless of the performance grading, were included. The grand means of the total durations of the periods when Fault Sum was displayed were 10.2 seconds in Elsie, and 11.8 seconds in Besi.

Table 4-3 summarizes various facets of operators' oculomotor behavior for the major Elsie and Besi display formats. The PFD was examined more frequently and for longer average durations in Besi than in Elsie, which explains the large percent PFD usages shown in the Besi bar graph in Figure 4-9. By contrast, Fault Sum and Fault Messages were examined less frequently, and for

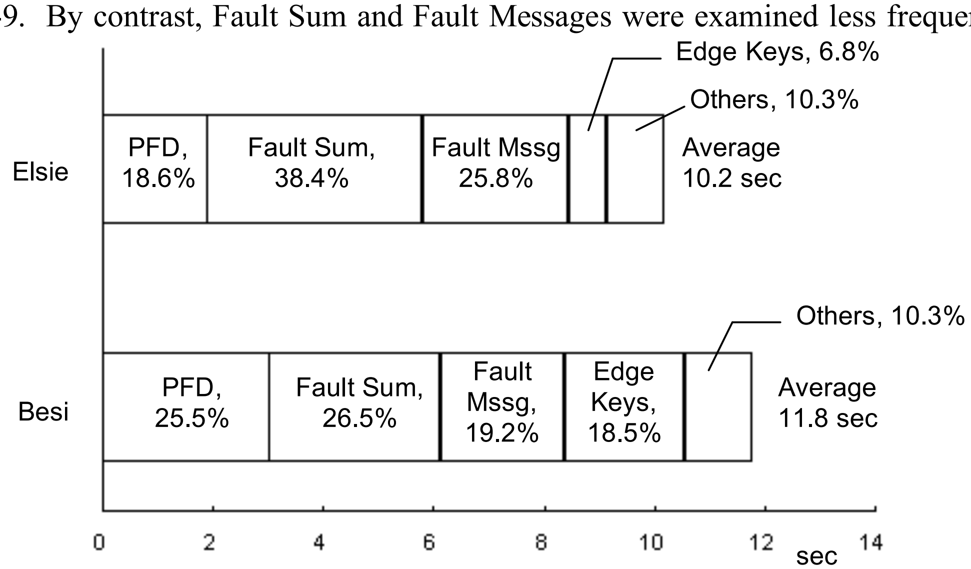


Figure 4-6. Display page usage during D1
for Fault Sum Sub-Phase

shorter average durations, in Besi than in Elsie. Note, also, that the edge keys received more frequent visits and longer average durations in Besi than in Elsie. These results suggest that Elsie's edge key label configuration afforded more efficient display navigation operations than Besi's matrix configuration, possibly because Elsie's edge key labels were spread over a wider area.

Table 4-3. Look Statistics during D1 when Fault Sum occupied the Main Area

Elsie			Besi		
Display	Number of looks per minute	Average duration per look [sec]	Display	Number of looks per minute	Average duration per look [sec]
PFD	12.8	0.87	PFD	16.2	0.94
Fault Sum	21.7	1.06	Fault Sum	18.8	0.85
Fault Mssg	9.6	1.61	Fault Mssg	8.5	1.35
Edge Keys	5.9	0.69	Edge Keys	7.0	1.59

Figure 4-7 illustrates the fixation transition probabilities among the most important ROI. The percentage values indicate the probabilities of transitioning in the shown direction from the origin ROI. Figure 4-7, combined with the numbers of looks per minute in Table 4-3, reveals that for Elsie, most fixation traffic occurred between adjacent display areas, that is, between Fault Sum and the PFD, and between Fault Sum and the fault messages/edge keys. Compared to these levels, traffic between the PFD and the nonadjacent (lower) ROIs (i.e., fault messages and the edge keys) was relatively light. In Besi, by contrast, traffic between the lower regions and

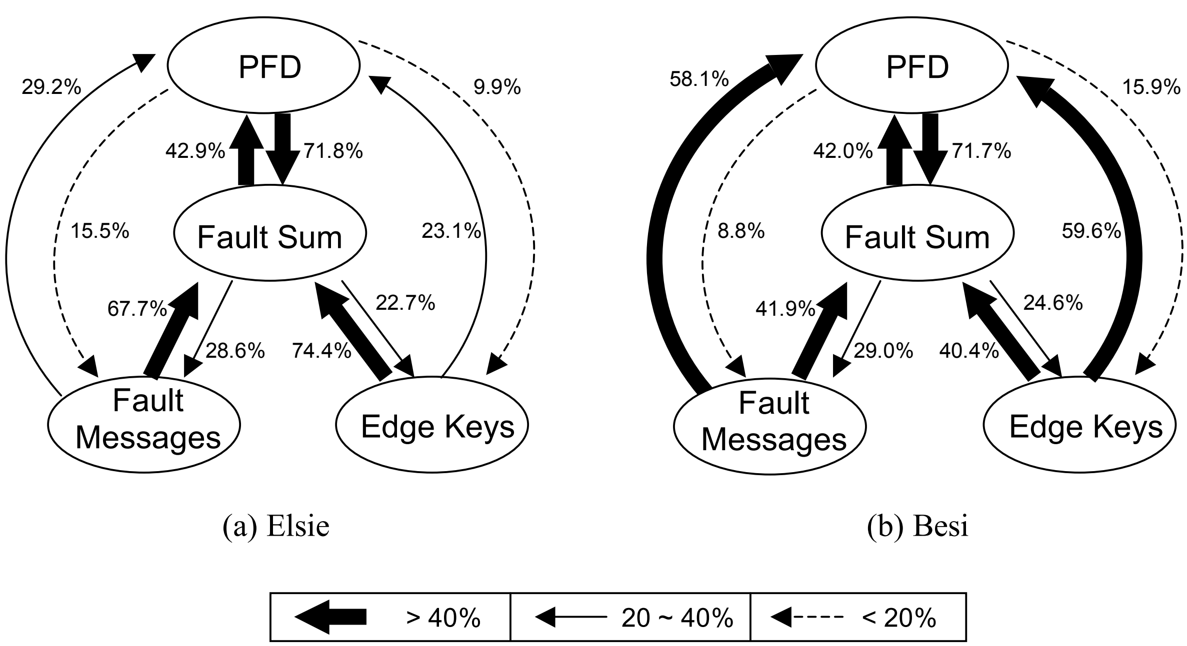


Figure 4-7. Transition Probabilities during D1
for Fault Sum Sub-Phase

the immediately adjacent (Fault Sum) display was considerably lighter, due to a pronounced *increase* in traffic between the lower regions and the (nonadjacent) PFD.

The EPS sub-phase

Figure 4-8 plots operators' time on display formats during D1 when the EPS displays (i.e., the EPS Sum or EPS Main for Elsie) or the EPS (Besi) were present in the Main Area (i.e., following a decision by the operator to replace the Fault Sum display with one of the EPS system summary displays). Data from all the paired trials (i.e., the scenarios #5 through #14), regardless of the performance grading, were included. The grand means of the total durations of the periods

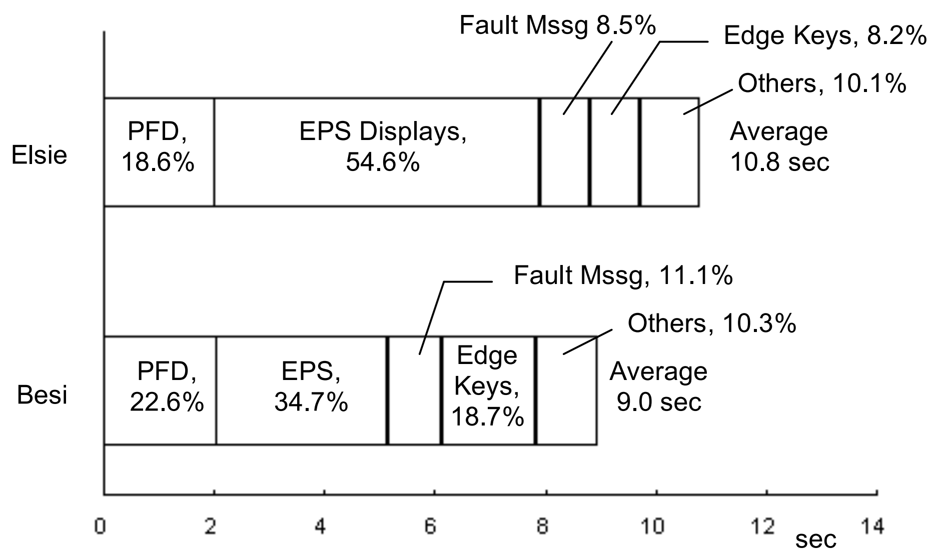


Figure 4-8. Display page usage during D1 for EPS sub phase

when the EPS-related displays were present were 10.8 seconds in Elsie and 9.0 seconds in Besi.

Table 4-4 summarizes eye movement behavior for the major displays used in Elsie and Besi. Replicating the pattern found in the Fault Sum Phase (Table 4-3), the PFD was again examined more frequently with Besi than with Elsie. Elsie's EPS Sum and EPS Main displays were also examined heavily during this period, consistent with the large (over 50%) usage fraction shown in the Elsie bar graph in Figure 4-8. Again, Besi's edge keys were fixated more frequently and for longer durations than Elsie's. However, unlike the previous period, and unlike Elsie during this period, operators working with the Besi interfaces did not need to use the edge keys to call up another other display in the main area. Thus, the exact reason why Besi's edge keys received more operator attention than Elsie's is unclear.

Table 4-4. Look Statistics during D1 when EPS-related displays were shown

Elsie			Besi		
Display	Number of looks per minute	Average duration per look [sec]	Display	Number of looks per minute	Average duration per look [sec]

PFD	12.6	0.85	PFD	14.3	0.95
EPS Sum & EPS Main	20.4	1.61	EPS	16.6	1.25
Fault Mssg	5.0	1.02	Fault Mssg	5.5	1.22
Edge Keys	4.7	1.05	Edge Keys	8.1	1.38

Figure 4-9 illustrates the fixation transition probabilities among the ROIs during the EPS period. The bar-graph information in Figure 4-8, together with the number of looks per minute data in Table 4-4, have already revealed that, in Elsie, IAEs to EPS Sum or EPS Main accounted for more than 50% of the total time. Clearly, the graphical information about system status and system functioning contained in these displays played a prominently role in fault diagnosis. Figure 4-12 reveals that some of the heaviest fixation traffic occurred between these EPS displays and the PFD. Traffic was considerable, if slightly more moderate, between the EPS displays and the lower regions, and moderate traffic also occurred from the lower regions to the PFD. Traffic from the PFD to the lower regions was again very light. In Besi, the major differences from these patterns came about because traffic got much heavier *from* the lower regions to the (nonadjacent) PFD, and also heavier (though to a lesser extent) in the reverse direction, from the PFD to the lower regions. This replicates the pattern in the earlier Fault Sum sub-phase.

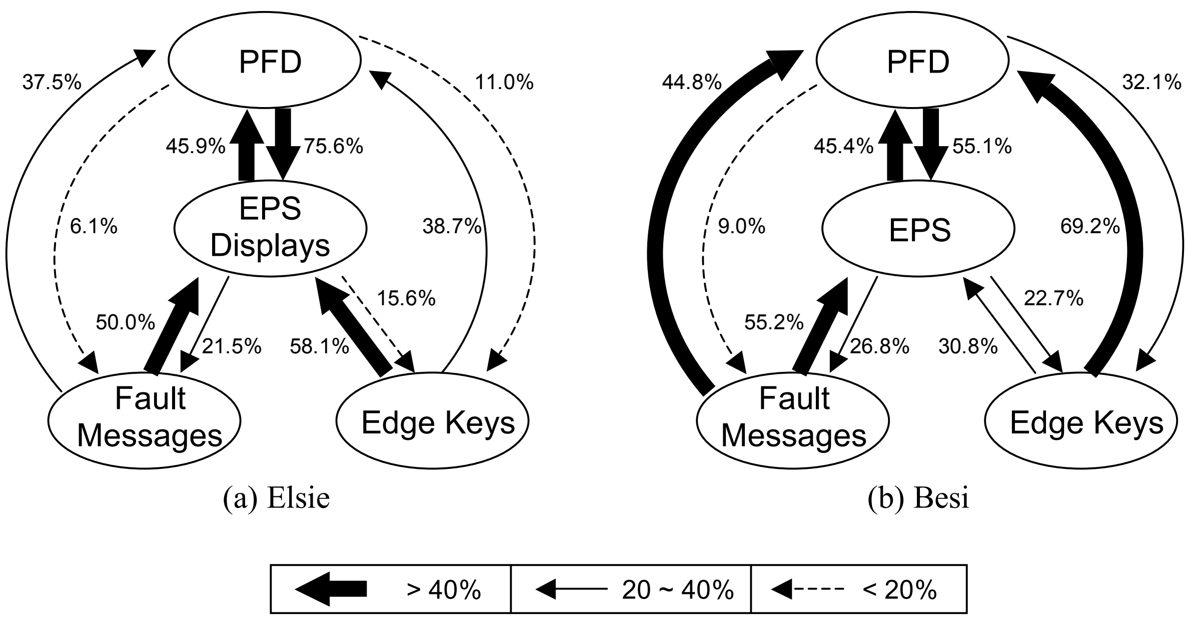


Figure 4-9. Transition Probabilities during D1 during the EPS sub phase.

The Fault Log Sub-Phase

Fault Log was never selected in Besi. Therefore, Figure 4-10 plots operators' display usages when Fault Log occupied the Main Area for Elsie only. The data from all the paired trials (i.e., the scenarios #5 through #14), regardless of the performance grading, were included. As shown in the Table, the mean duration of this period was 15.8 seconds. Almost 60% of this time was occupied by looking at Fault Log itself.

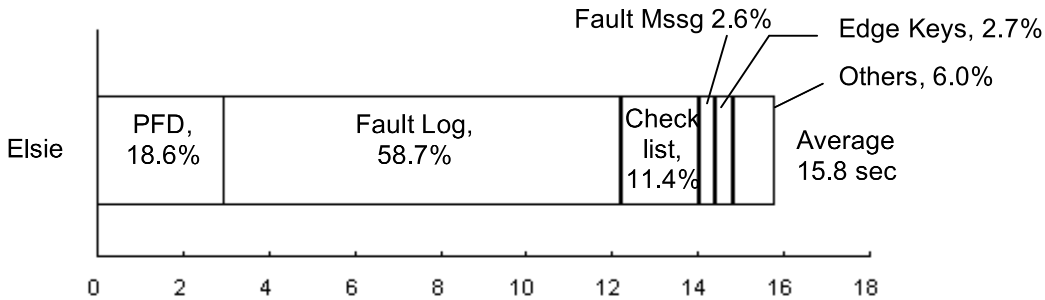


Figure 4-10. Display page usage during D1
During Fault Log Sub-Phase (Elsie)

Table 4-5 lists the summary oculomotor statistics for the major Elsie displays during this period. Although fixation frequency was not noticeably different between the PFD and Fault Log, reading the densely packed fault messages produced an average fixation duration, 2.38 sec, that was by far the longest of all analyzed durations in Tables 4-3, 4-4, and 4-5. This explains the large discrepancy in amount of time spent on the PFD compared to Fault Log (18.6% versus 58.7%) in Figure 4-10.

Table 4-5. Look Statistics during D1 when Fault Log was shown

Elsie		
Display	Number of looks per minute	Average duration per look [sec]
PFD	13.0	0.87
Fault Log	17.1	2.38
Checklist	5.0	1.36

Figure 4-11 plots the fixation transition probabilities among the major Elsie displays. Figures 4-11, combined with the information in Figure 4-10 and the number of looks per minute data in Table 4-5, indicate heavy eye traffic between the PFD and the Fault Log. However, operators also started to turn their attention to the checklist area, presumably to start the process of associating fault log messages with the checklist titles in the EPV menus, in preparation for selecting the appropriate checklist from the EPV menu.

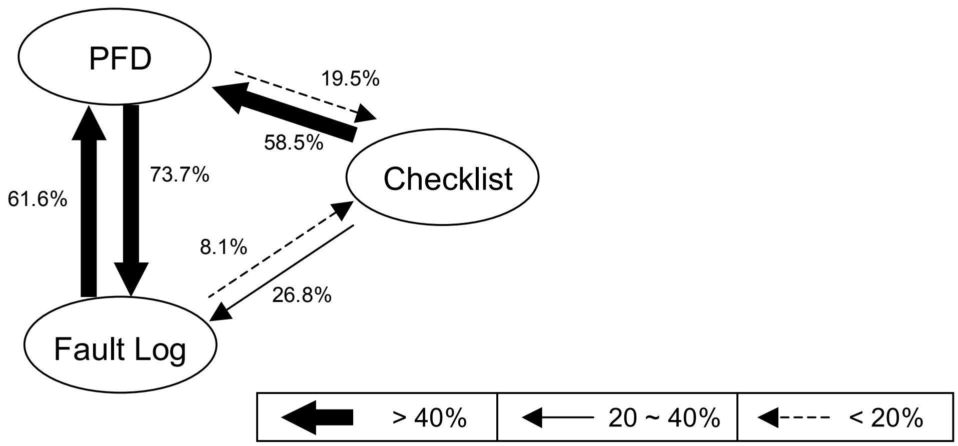


Figure 4-11. Transition Probabilities during D1 in Elsie
During Fault Log Sub-Phase

Re-visitations

A fundamental difference between the PFD-based task and the fault management task is that the PFD-based task was based on detection of a periodic, and somewhat temporally unpredictable, color change, so it is not surprising that operators elected to glance at the PFD on a regular basis. The fault management task, on the other hand, involved multiple discrete sources of information whose content was largely fixed (static) from the outset of the malfunction. As we discussed, the static nature of the fault management information meant that, in principle, operators could have elected to process all diagnosis-relevant information from a particular ROI during a single IAE. An alternative strategy would be to acquire information from the relevant displays in a more incremental fashion, across several distinct IAEs, separated by IAEs to other regions of interest.

To distinguish between these possibilities, we counted the number of trials that involved two or more distinct IAEs (i.e., at least one revisit) to three malfunction-relevant ROI's: Fault Sum, the lower C&W system Message Area, and the EPS displays. For Fault Sum, fully 84% of the trials involved at least two IAE's (one revisit), and 68% involved more than two IAEs (i.e., two or more revisits after the initial IAE). Very similar percentages were obtained for both the continuously visible Message Area and the EPS displays.

These results obviously favor an incremental information acquisition strategy, separated by IAE's to other regions of interest, as opposed to an exhaustive "all-in-one" strategy. In addition, note that across all D1 phases, there were always two ROIs containing fault-related information: either Fault Sum, the EPS displays, or Fault Log in the Main Display Area, and the continuously viewable subset of C&W fault messages in the lower Fault Message Area. As revealed in Figures 4-7 and 4-9, there was considerable eye-movement traffic between these regions, particularly in Elsie (for example, more than half of the traffic from the Fault Message Area went directly to Fault Sum; see Figure 4-7). Together with the revisiting results, the pattern suggests that, not only did operators adopt an incremental information acquisition strategy, but they supported their proximal-cause determination process by cross-checking the information gathered from an individual IAE to one fault-management relevant display against the information garnered from an IAE to the other. This cross-checking strategy is particularly noteworthy, given that the format of the information across the ROIs was so distinct, being text-based in one region (Fault Message Area), and more graphical (e.g., EPS Sum) in the other.

5. Discussion

In our earlier SHFE report (Hayashi et al., 2007), we assessed two concepts for fault management operational concepts, Elsie and Besi, using a traditional suite of human factors performance measures, such as completion (response) time for the individual phases of the fault management task, error rates, and subjective assessments of workload and usability. These measures were sufficient to reveal the general impact of automation on fault management, including the fact that, compared to Elsie, Besi reduced the duration of the D1 (diagnosis) phase by approximately 50%, and that fault management activities impaired performance on the PFD task more with Elsie than Besi. These measures might be sufficient to allow program managers to determine whether targeted operational concepts and associated crew interfaces meet high-level requirements for accuracy and workload. However, they provide little explicit guidance to the operational concept design process. Eye movement analyses allow us to fill in valuable gaps in our understanding of operator's information acquisition patterns and strategies, and provide much more explicit guidance for operational concept and user interface design.

5.1. Impact of Automation: Elsie versus Besi

As noted, compared to Elsie, working malfunctions with the Besi interfaces reduced the duration of the diagnosis (D1) phase by approximately 50% (an operationally significant 19 sec). Display usage patterns revealed that this reduction was largely a consequence of the fact that Elsie involved much more extensive processing of text information within text-rich display formats. Of the 19 sec Elsie penalty, fully half (9.4 sec) was accounted for by reading C&W fault messages on the Fault Log display. Moreover, in the period immediately following the initial diagnosis stage (Stage C1), navigating to and bringing up the appropriate procedure checklist consumed a further 20 sec of operator time in Elsie, compared to less than a second in Besi, where the operation was automated. Once again, the majority of the time penalty required to complete Stage C1 in Elsie was accounted for by processing text within text-rich displays (specifically, the menus of systems and checklist titles on the EPV and the list of fault messages on the Fault Log).

Taken together, these findings highlight an important conclusion: Since processing cluttered text-rich displays (see, e.g., Figure 2-9) confers a major time penalty on human operators, *designers of fault management operational concepts should minimize the need to present and process information on such display formats whenever possible.*

Fortunately, the development of a new generation of crewed vehicles provides many opportunities to do so. An advanced C&W system with proximate-cause diagnosis capabilities obviously reduces or eliminates the need to process fault messages on a fault log display. Other opportunities may be less obvious. Suppose, for example, vehicle weight and software development, test, and validation considerations tip the scale in favor of a less capable (but less computationally intensive) Elsie-like fault management system, as opposed to a system with Besi's automation. It would still be quite feasible to reduce the need to process lists of text elements on EPV menus with direct retrieval of the procedure checklists via fault log messages.

That is, the Fault Log page could include a "focus" bar to enable cursor focus on individual fault messages. Following the lead of the B777, participants could scroll the focus bar to whatever fault message they determined to be the proximal cause, select that choice with a button press, and retrieve the corresponding checklist in the EPV area automatically. Our results suggest that this relatively modest modification to the Elsie operational concept could result in as much as 20 seconds of savings, largely because it would eliminate the need to bring up a checklist via EPV menus. Since the functional links between the C&W database and the EPV database necessary to produce this capability could be programmed in advance, the performance benefit would be achieved with little additional demand on onboard computational resources.

This line of thought brings up an important ancillary point. EPV development for Orion could proceed with a mindset that the EPV would operate largely independently of other onboard software systems, with the major exception of vehicle commanding channels. This "stand-alone" perspective would naturally produce the "menu-based" concept for navigating to and retrieving EPV checklists; indeed, with a functionally self-contained operational concept, an alternative method for checklist navigation is not immediately obvious. However, a much different concept for checklist retrieval comes about from considering the C&W system and the EPV as interacting components supporting the same overarching task (real-time fault management). The present results provide a strong empirical case that the EPV and the C&W system should proceed from an integrated (task-oriented) perspective. Our results provide empirical "benchmarks" against which to assess the operational benefits that arise from building functional linkages between the software for these traditionally isolated systems.

Such linkages would also help mitigate the risk that transitioning to the EPV from paper may actually reduce fault management efficiency in particular circumstances. Paper provides a degree of flexibility, particularly for accessing checklists, that a standalone EPV may be hard pressed to match. For example, in the shuttle cockpits, the checklists for working the most likely propulsion system malfunctions during ascent are available, not only within flight data file booklets, but also on cue cards within easy reach of the flight crew. Compared to the brief time required to grab and start reading a cue card, our results suggest that the time needed to navigate through EPV menus to call up these procedures might translate into a considerable performance decrement. A more user-centered (integrated) approach to the design and evaluation of fault management systems can reveal opportunities to ameliorate these potential drawbacks, by A) empirically determining those activities that suffer from the electronic transition, and B) suggesting opportunities to ameliorate them through targeted automation.

Even if the choice was made to implement a standalone or "bare-bones" C&W system, with no functional links to the EPV, our results still suggest design options that could have considerable utility. For example, designers could exploit preexisting failure modes, effects, and criticality analyses to organize the list of fault messages on the fault log, perceptually segregating the most common "parent" fault messages from the most common "children" (for example, grouping the children messages and indenting them under the "parent" messages, using different font sizes for parents and children, or some combination), enabling "at a glance" perceptual segregation of the subset of generated fault messages that form the best proximal-cause candidates. Alternatively, with just a modest additional software integration effort, text-processing demands could be reduced by visually linking the information on the Fault Log page with the checklist menus through "yoked" focus bars. That is, assuming the text line (checklist title) with current cursor focus is indicated on the EPV menu with a focus bar, the matching entry (text line) within the fault log page (if present) would be highlighted with a corresponding focus bar, and move in concert with each other while the operator was scrolling. That way, the critical connection would be established perceptually, making it easier to cross-check (integrate) the information on the two displays with a minimum of extraneous text processing.

5.2. Display Real Estate Requirements

Operators were trained to incorporate both caution and warning (fault messages) information and information from the more graphical displays (such as the Fault Sum, EPS Sum, EPS Main, and Besi's EPS displays) into their proximal-cause diagnosis strategies. However, there was considerable freedom over just how to acquire and integrate the information from the various sources. One strategic issue is particularly germane because of possible severe restrictions on display real estate available to support fault management operations. Our operators acquired information from the relevant displays in a temporally distributed manner with multiple individual information extraction episodes interleaved with episodes from other display areas. This suggests that designers should ensure that their caution and warning interfaces allow for simultaneous viewing of graphical/system summary displays (i.e., the Fault Sum and the EPS-related displays) and fault message displays (such as the Fault Log display and the Fault Message area), as opposed to forcing the two formats to share the same real estate. By extension, it may have been quite detrimental in Elsie to force operators to time-share the Fault Log display with the more graphical EPS sum and Fault Sum displays, preventing simultaneous viewing of these display formats. We speculate that if the interface was changed to allow simultaneous presentation of Fault Log and the more graphical display formats, thus enabling an interleaving strategy across all phases of D1, the duration of the fault diagnosis stage would have been reduced. This change could be easily implemented, by simply populating the EPV area with the C&W fault log as soon as the fault messages are generated.

From a more theoretical perspective, the evidence for repeated sampling of the same display areas during the diagnosis stage (in particular, repeated viewing of the message and EPS summary displays during the middle segment of D1) suggests that information acquisition during diagnosis is not an "all-or-none" phenomenon. Factors that might be contributing to this are:

1) Fault Diagnosis is a "Random-Walk". The process of converging on a "proximal cause" decision is essentially an exercise in deciding on (selecting) one from a set of candidates defined by the fault messages generated by the caution and warning system, supported by other forms of fault management information. One way of modeling the diagnostic process would be to consider each sequential information sampling activity (fixation) as providing incremental evidence for or against each individual candidate in a hypothesis set. Suppose each candidate is associated with its own "evidence counter," which would be either incremented when the information extracted from the current fixation provides evidence for that candidate, or decremented when the fixation supports another candidate. In this framework, the value of the counter for each hypothesis would follow a random walk over time until the count (strength of evidence) for one candidate exceeded the sum of the counts for the alternatives by some critical ratio. In this framework, re-fixating previously sampled regions would continue until the critical ratio was exceeded, at which point the operator would move on to the next stage in the fault management task (selecting the correct checklist). As long as the evidence accumulation process

is noisy and incremental, repeated fixations on the same displays would be necessary to drive the "winning" hypothesis past threshold.

2) Memory decay. We know from the eye movement analyses in Table 3-1 that participants implemented a more-or-less continuous sampling strategy vis a vis the PFD, as revealed by the constant count of PFD fixations per minute during the active fault management period (a rate which was, however, higher for Besi than for Elsie; see below). Importantly, each distinct check of the PFD interrupted fault management activities. It is quite possible that the information acquired from the fixations preceding each PFD interruption decayed during the "interruption period, forcing operators to re-acquire previously-processed information following the interruption. Again, this would result in an increase in the incidence of re-fixation behavior on previously sampled displays and display sub-regions.

3) Multi-tasking strategies. Last but not least, the eye movement analyses revealed that the multi-tasking performance benefits associated with the Besi interface reflected a simple continuous multi-tasking strategy whereby operators interrupted fault management task activities to sample the PFD more frequently while working malfunctions with Besi than with Elsie. Importantly, providing a greater level of automated support for fault management enabled more than just shorter and more accurate fault management operations per se. It also engendered more willingness on the part of participants to spread their attention between fault management activities and other simultaneous operational demands, particularly from areas further away from the PFD. It is quite noteworthy that this benefit was not confined to the diagnosis phase itself, and was therefore not narrowly associated with the automated proximal-cause capability that came with Besi. Instead, we speculate that all of the different sources of automated aid provided by Besi combined to influence a high-level scheduling algorithm to adjust one of its parameters to allow for a higher interruption frequency.

The pattern of transition probabilities in Besi and Elsie in the first two sub-phases of the diagnosis phase (when Fault Sum was up [Figure 4-7], and when the EPS display(s) were up [Figure 4-9]), provide further interesting clues to the impact of Elsie on operator's attentional strategies. In both figures, the most notable change in traffic between the various regions of interest was spatial. In Elsie, traffic patterns were highly influenced by proximity; operators were most likely to transition to a region of interest immediately adjacent to the region sampled on the previous fixation. For example, there was relatively strong traffic between the lowest regions (the Fault Message Area and Edge Keys) and the immediately adjacent region (either Fault Sum or the EPS display[s]). There was relatively light traffic between the lower regions and the PFD. In Besi, by contrast, the influence of physical proximity was considerably weaker, as evidenced by the large increase in traffic between the lower regions and the PFD.

There are two possible accounts for this change. It may be that programming large distance saccades imposes more of a cognitive load than short-distance saccades (Ballard, Hayhoe, & Pelz, 1995). Since Besi automated more activities than Elsie, the Besi concept "freed up" cognitive resources to support more long-distance saccades. Alternatively, the increase in long-distance traffic in Besi may have been a more task-related effect. Since, in Elsie, operators were more cognitively involved on the fault management task, top-down attentional settings may have been imposed that assigned the central displays (that were all fault-management information displays) a higher saliency than they were accorded in Besi (Folk, Remington, and Johnston, 1992). Operators were thus more likely to visit these central displays than they were in Besi, which would masquerade as a physical proximity effect.

6. Conclusions

In this follow-up to Hayashi, et al. (2007), we used the results of more fine-grained analyses of participants' eye movements to derive guidelines and suggestions for interface design and human-machine functional allocation for fault management systems on next-generation spacecraft. Due to weight and schedule pressures, it is at present unclear how much automation will be built into the Orion Block 1 fault management system, so it is unclear how closely the fault management operations concept will resemble the highly automated Besi as compared to more manual Elsie. One of the biggest question marks is whether Orion's C&W system will incorporate an automated diagnostic capability. Eye movement analyses provided further insight into the impacts of incorporating such a capability, some of which were not obvious from the earlier report. In addition, however, the analyses revealed some important performance benefits from other forms of Besi automation, particularly for information retrieval and display. These forms are less computationally complex, and would generate far less development and operational risk, than proximal-cause software. They could be added to an Elsie-style fault management system to create an "Elsie-Besi hybrid" that achieves a substantial fraction of the Besi performance enhancements for a relatively modest investment in software development.

In addition to providing a more sensitive set of tools with which to evaluate the impact of automation, the eye movement analyses identified display formats and concepts that were particularly time consuming for operators to process. Uncovering and quantifying these inefficiencies enabled us to make specific recommendations for the fault management operational concept. Once again, many of these redesign ideas could be achieved with little or no additional demands on the onboard computational resources.

One point is clear. If caution and warning systems development proceeds largely independently of the development of the EPV and its supporting software, the operations community will miss a major opportunity to improve the efficiency of fault management activities. In the past, such stove piping of the two fault management activities was a natural product of the fact that, because procedures were paper-based, there were no opportunities for functional integration. Today, when the databases supporting the cockpit displays for both systems are going electronic, operationally significant improvements can be achieved by building in functional links between the caution and warning system and the EPV.

7. References

Ballard, D. H., Hayhoe, M. M., & Pelz, J. B. (1995). Memory representations in natural tasks. *Journal of Cognitive Neuroscience*, *7*, 66-80.

Droll, J. A., & Hayhoe, M. M. (2007). Trade-offs between gaze and working memory use. *Journal of Experimental Psychology: Human Perception and Performance*, *33*, 1352-1365

Droll, J. A., Hayhoe, M. M., Treisch, J., & Sullivan, B. T. (2005). Task demands control acquisition and storage of visual information. *Journal of Experimental Psychology: Human Perception and Performance*, *31*, 1416-1438.

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 1030-1044.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In P. A. Hancock & N. Meshkati (Eds.), *Human Mental Workload* (pp. 139-183). Amsterdam: North Holland Press.

Hayashi, M., McCann, R. S., Beutter, B. R., Spirkovska, L., Poll, S., Sweet, A., et al. (2007). *Human Factors Evaluation of Caution and Warning Interface Concepts for Project Constellation Vehicles*. SHFE NASA Report.

Huntley, M. S. (1993). *Flight Technical Error for Category B Non-Precision Approaches and Missed Approaches Using Non-Differential GPS for Course Guidance* (No. DOT-VNTSC-FAA-93-17). Cambridge, MA: US Dept. of Transportation Volpe National Transportation Systems Center.

McCann, R., Beutter, B. R., Matessa, M., McCandless, J. W., Spirkovska, L., Liston, D., Hayashi, M., Ravinder, U., Elkins, S., Renema, F., Lawrence, R., & Hamilton, A. (2006). *Description and evaluation of a real-time fault management concept for next-generation space vehicles*. SHFE NASA report.

McCandless, J., Hilty, B., & McCann, R. S. (2005). New displays for the space shuttle cockpit. *Ergonomics in Design*, *13*, 15-20.

Poll, S. A. Patterson-Hine, A., Camisa, J., Garcia, D, Hall, D., Lee, C., Mengshoel, O. J., Neukom, C., Nishikawa, D., Ossenfort, J., Sweet, A., Yentus, S., Roychoudhury, I., Daigle, M., Biswas, G., and Koutsoukos, X. (2007). Advanced diagnostics and prognostics testbed. In *Proceedings of the 18th International Workshop on Principles of Diagnosis (DX-07),* 178-185

Roscoe, A. H. (1984). *Assessing pilot workload in flight*. In Advisory Group for Aerospace Research & Development Conference Proceedings N0. 373: Flight Test Techniques. Neuilly-sur-Seine, France: NATO.

8. Appendix A: Eye-movement analyses following C1

8.1. Percent Display Usages during D2

Figure 8-1 shows the percent usages of the displays during D2 of the single-malfunction trials. As in the previous bar charts, only the data from the trials graded *Correct* were included in the bar charts. The grand average durations were 41 sec for Elsie and 38 sec for Besi.

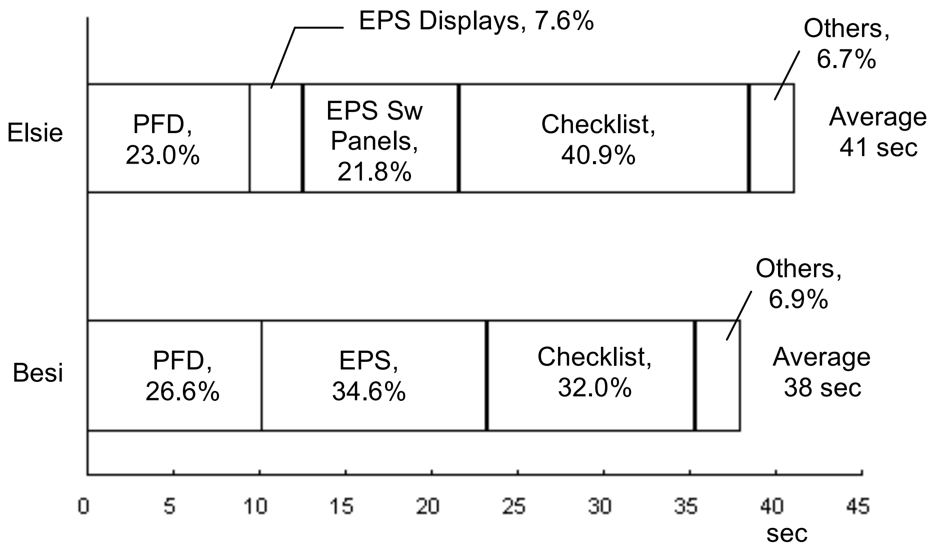


Figure 8-1. Display page usage during D2

The Elsie's EPS Displays ROI was comprised of the EPS Main (5.8%) and the EPL Loads (1.8%). The EPS Switch Panels contained the EPS Dist Switch Panel (12.5%) and the EPS Loads Switch Panel (9.2%). The *Others* of Elsie included the Fault Log (0.2%) and the Message (0.2%), and the *Others* of Besi the Fault Sum (0.2%) and the Message (2.3%). The remaining of the *Others* categories was saccades among ROI or noise.

The D2 process was basically led by the checklist instructions. Indeed, the Checklist ROI accounted for large portion of the total D2 durations in both Elsie and Besi. The EPS Displays (Elsie), the EPS Switch Panels (Elsie), and the EPS (Besi) displays were also used as required by the checklist steps. Interestingly, Besi contained slightly more total PFD look than Elsie, despite that the average total D2 durations were shorter for Besi than Elsie. This point will be revisited in Section 3.2.2.

8.2. Percent Display Usages during C2

Figure 8-2 shows the percent usages of the displays during C2 of the single-malfunction trials. As in the previous bar charts, only the data from the trials graded *Correct* were included in the bar charts. The grand average durations were 14 sec for Elsie and 1.6 sec for Besi.

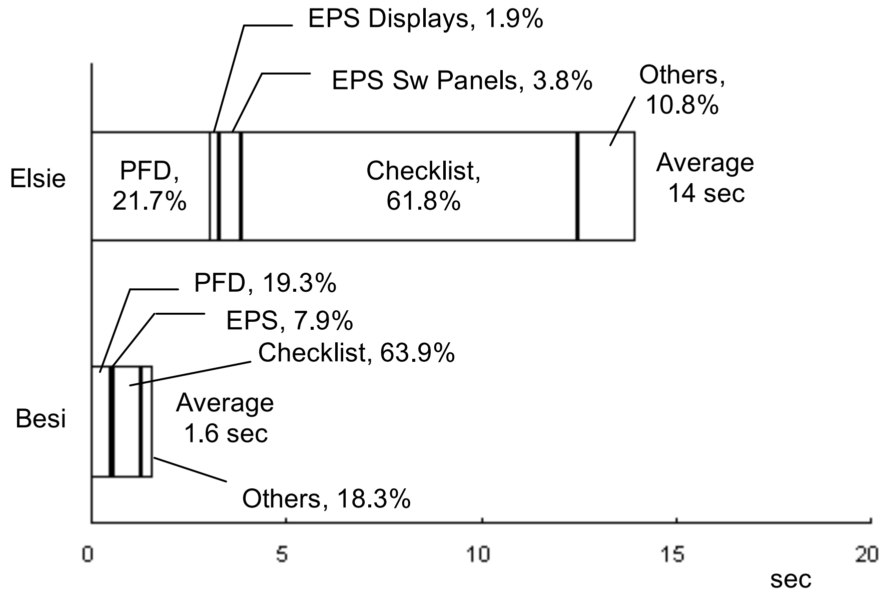


Figure 8-2. Display page usage during C2

The EPS Displays ROI encompassed the EPS Main (1.8%) and the EPS Loads (0.1%). The EPS Switch Panels contained the EPS Dist Switch Panel only (i.e., no EPS Loads Switch Panel). The *Others* category of Elsie included the Fault Log (2.4%) and the Message (1.3%). The *Others* of Besi consisted of the Message (0.7%). The remaining of the *Others* categories were saccades among the ROI or noise.

As expected, the Elsie graph shows that the largest portion of the total C2 time was spent on the Checklist ROI.

8.3. Percent Display Usages during R

Figure 8-3 shows the percent usages of the displays during R of the single-malfunction trials. As in the previous bar charts, only the data from the trials graded *Correct* were included in the bar charts. The grand average durations were 63 sec for Elsie and 67 sec for Besi.

The EPS Switch Panels contained the EPS Dist Switch Panel (6.0%) and the EPS Loads Switch Panel (33.6%). The *Others* of Elsie included the EPS Main (0.2%) and the Message (0.2%). The remaining *Others* of Elsie and all *Others* of Besi were saccades between ROIs and noise.

Like the D2 process, the R process was also basically led by the checklist instructions. As a result, the Checklist ROI accounted for large portion of the total R durations in both Elsie and Besi. The EPS Switch Panels (Elsie) or the EPS (Besi) displays were used as the checklist required the recovery switch throws. Again, as in the D2 phase, the participants had looked at PFD for more time in Besi than in Elsie. This time, it had caused longer average total R durations for Besi than for Elsie. See the Section 3.2.2 for further analysis.

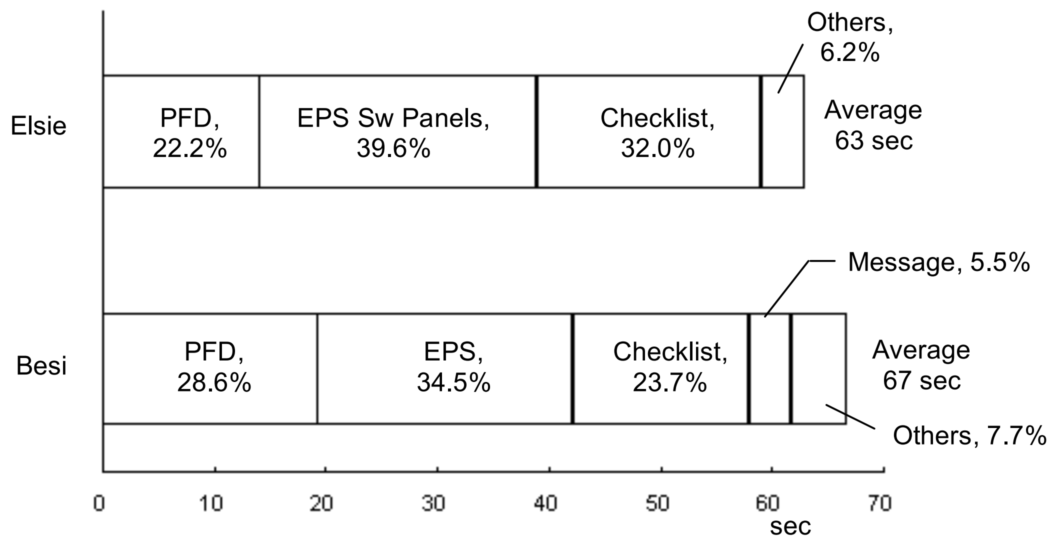


Figure 8-3. Display page usage during R

9. Appendix B: PFD Look Statistics during D2 and R

In Figures 8-1 and 8-3, the bar charts showed that more of the total D2 and R durations was spent monitoring the PFD in Besi than in Elsie. To further investigate this result, the three PFD-look statistics during D2 and R, respectively, were computed and summarized in Table 9-1. Only the data from the single-malfunction trials (#5 through #10) were included. Note that the values of the % time of PFD look are slightly different from those shown in the bar charts, because the values in Table 4-2 include the data from the trials categorized as *Good* and *Failed* as well (the bar charges included the only data from the single-malfunction trials graded *Correct*). This maximized the number of data subjected to the paired t-tests to be performed next. Again, the focus was how their PFD-monitoring performance was influenced by fault management concept, regardless of the specific fault management task (i.e., phase) and malfunction-handling accuracy.

Table 9-1. PFD Look Statistics during Phases D2 and R

(a) During D2

Scenario Difficulty x Concept	N	% time of PFD look	Number of PFD looks per minute	Average PFD look duration [sec]
Single-mal Elsie	24	20.9 %	15.0	0.84
Single-mal Besi	24	24.9 %	19.6	0.80

(b) During R

Scenario Difficulty x Concept	N	% time of PFD look	Number of PFD looks per minute	Average PFD look duration [sec]
Single-mal Elsie	23	22.3 %	16.5	0.84
Single-mal Besi	23	27.3 %	17.4	0.96

In the D2 phase, the paired t-test results indicated that the participants looked at the PFD significantly more frequently in Besi than in Elsie, $t(23) = 3.18$, $p < 0.01$. There was no statistically significant difference in the percent time of PFD look and the average PFD look durations.

In the R phase, the results also showed that operators looked at the PFD for a significantly larger proportion of the total duration with Besi than with Elsie, $t(22) = 2.29$, $p = 0.03$, and the average PFD look durations were marginally longer with Besi than with Elsie, $t(22) = 1.86$, $p = 0.077$. No statistical significance was found for the number of PFD looks per minute. Note that the total number of data included, *N*, was 23 in the R, because one participant did not finish the R phase of scenario #7 in Besi. Thus, this trial and the paired trial (#8 in Elsie) of this participant were not included in the t-tests.